

Universität Duisburg-Essen  
Fakultät für Bildungswissenschaften  
Lehrstuhl für Lehr-Lernpsychologie

# **Analytisches Problemlösen: Validität und Potenzialnutzung**

Dissertation zur Erlangung des Grades Dr. phil.  
vorgelegt von Dipl.-Psych. Jens Fleischer  
geboren am 02.10.1976 in Böblingen

Erstgutachter: Prof. Dr. Dr. Detlev Leutner, Universität Duisburg-Essen

Zweitgutachter: Prof. Dr. Joachim Wirth, Ruhr-Universität Bochum

Tag der mündlichen Prüfung: 23. August 2017

## Danksagung

Zunächst möchte ich mich ganz herzlich bei Herrn Prof. Dr. Leutner für die gute Betreuung sowie die wertvollen Anregungen und Ratschläge bei der Umsetzung dieser Arbeit bedanken.

Mein ganz herzlicher Dank gilt auch Herrn Prof. Dr. Wirth für die Bereitschaft, diese Arbeit als Zweitgutachter zu betreuen und für die konstruktive Kritik, von der diese Arbeit sehr profitiert hat.

Ein besonderer Dank geht an Christian (Spoden) für die sehr gute Zusammenarbeit und dafür, dass er stets bei methodischen Problemen beratend zur Seite stand, auch wenn es mal kniffliger wurde. Ein großer Dank geht auch an Florian (Buchwald) für den hilfreichen Erfahrungsaustausch und die sehr gute Zusammenarbeit im DFG-Schwerpunktprogramm „Kompetenzmodelle“.

Ebenfalls bedanken möchte ich mich bei allen studentischen Hilfskräften, die diese Arbeit nicht unwesentlich unterstützt haben.

Zu guter Letzt und ganz besonders: Dank an Marie für alles, und an Emil für die gute Laune!

# Inhaltsverzeichnis

Liste der Veröffentlichungen von Teilen der Arbeit .....	5
Tabellenverzeichnis.....	6
Abbildungsverzeichnis .....	8
Zusammenfassung .....	9
Abstract.....	10
<b>1 Allgemeine Einführung und Struktur der Arbeit .....</b>	<b>11</b>
1.1 Probleme und Problemlösen .....	11
1.2 Problemlösen als Prozess.....	14
1.3 Problemlösen als Kompetenz .....	17
1.4 Struktur und Forschungsfragen .....	31
1.5 Literatur .....	34
<b>2 Studie 1: Faktorielle Validität (Dimensionalität) der Modellierung analytischer Problemlösekompetenz – Die Bedeutung unterschiedlicher Arten von Problemstellungen .....</b>	<b>48</b>
2.1 Einleitung .....	48
2.2 Theoretischer Hintergrund .....	49
2.3 Fragestellungen und Hypothesen .....	53
2.4 Methode .....	54
2.5 Ergebnisse .....	59
2.6 Diskussion .....	63
2.7 Literatur .....	68
<b>3 Studie 2: Diskriminante Validität der Modellierung analytischer Problemlösekompetenz – Abgrenzung von fluiden Fähigkeiten .....</b>	<b>73</b>
3.1 Einleitung .....	73
3.2 Theoretischer Hintergrund .....	74
3.3 Fragestellungen und Hypothesen .....	79
3.4 Methode .....	80
3.5 Ergebnisse .....	86
3.6 Diskussion .....	93
3.7 Literatur .....	99

<b>4 Studie 3: Prognostische und inkrementelle Validität der Modellierung analytischer Problemlösekompetenz – Vorhersage zukünftiger fachbezogener Kompetenzen .....</b>	<b>107</b>
4.1 Einleitung .....	107
4.2 Theoretischer Hintergrund .....	108
4.3 Fragestellungen und Hypothesen .....	111
4.4 Methode .....	113
4.5 Ergebnisse .....	115
4.6 Diskussion .....	121
4.7 Literatur .....	128
 <b>5 Studie 4: Effekte der kontextuellen Einkleidung von Testaufgaben auf die Schülerleistungen im analytischen Problemlösen und in der Mathematik .....</b>	 <b>134</b>
5.1 Einleitung .....	134
5.2 Theoretischer Hintergrund .....	135
5.3 Fragestellungen und Hypothesen .....	140
5.4 Methode .....	142
5.5 Ergebnisse .....	145
5.6 Diskussion .....	151
5.7 Literatur .....	157
 <b>6 Gesamtdiskussion .....</b>	 <b>163</b>
6.1 Zentrale Ergebnisse.....	164
6.2 Theoretischer Ertrag .....	167
6.3 Praktischer Ertrag.....	171
6.4 Limitationen und Ausblick .....	173
6.5 Literatur .....	177
 <b>7 Anhang .....</b>	 <b>183</b>
Anhang A1: Testheftcover und einführender Text in der Kontextbedingung „Mathematik“ .....	183
Anhang A2: Einführendes Item in der Kontextbedingung „Mathematik“.....	184
Anhang B1: Testheftcover und einführender Text in der Kontextbedingung „Problemlösen“ .....	185
Anhang B2: Einführendes Item in der Kontextbedingung „Problemlösen“ .....	186
Anhang C: Lebenslauf .....	187

## Liste der Veröffentlichungen von Teilen der Arbeit

- Leutner, D., Fleischer, J. & Wirth, K. (2006). Problemlösekompetenz als Prädiktor für zukünftige Kompetenz in Mathematik und in den Naturwissenschaften. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 119–137). Münster: Waxmann.
- Fleischer, J., Wirth, J. & Leutner, D. (2014). Effekte der kontextuellen Einkleidung von Testaufgaben auf die Schülerleistungen im analytischen Problemlösen und in der Mathematik. *Zeitschrift für Pädagogische Psychologie*, 28, 217–227. doi: 10.1024/1010-0652/a000135
- Fleischer, J., Buchwald, F., Wirth, J., Rumann, S. & Leutner, D. (2017). Analytical problem solving: Potentials and manifestations. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 33–45). Paris: OECD Publishing. doi: 10.1787/20769679

## Tabellenverzeichnis

<i>Tabelle 2.1</i>	Ziele, kognitive Prozesse und potenziell schwierigkeiterzeugende Merkmale der drei Arten von Problemstellungen aus PISA 2003.....	51
<i>Tabelle 2.2</i>	Reliabilitäten, Interkorrelationen und Varianzen ( $\sigma^2$ ) der latenten Dimensionen des dreidimensionalen Modells der analytischen Problemlösekompetenz sowie Infit-Werte der zugehörigen Testitems.....	59
<i>Tabelle 2.3</i>	Modellpassung für das eindimensionale und das dreidimensionale Modell der analytischen Problemlösekompetenz.....	60
<i>Tabelle 2.4</i>	Intraklassenkorrelationen (ICC) und Mittelwertdifferenzen der Personenparameter des dreidimensionalen Modells der analytischen Problemlösekompetenz nach Bildungsgang, Familiensprache und Geschlecht sowie Korrelationen mit fachbezogenen Kompetenzen und fluiden Fähigkeiten (95%-Konfidenzintervalle in eckigen Klammern).....	62
<i>Tabelle 3.1</i>	Reliabilitäten (EAP-Schätzungen), Varianzen ( $\sigma^2$ ) und Infit-Werte der Dimensionen der analytischen Problemlösekompetenz und der fluiden Fähigkeiten.....	82
<i>Tabelle 3.2</i>	Anzahl geschätzter Parameter sowie Angaben zur Modellpassung des Nested-Faktor-Modells, des erweiterten Standardmodells sowie des Einfaktormodells.....	87
<i>Tabelle 3.3</i>	Intraklassenkorrelationen (ICC) und Mittelwertdifferenzen der Faktorwerte des erweiterten Standardmodells und des Nested-Faktor-Modells nach Bildungsgang, Familiensprache und Geschlecht sowie Korrelationen mit fachbezogenen Kompetenzen (95%-Konfidenzintervalle in eckigen Klammern).....	92

<i>Tabelle 4.1</i>	Regressionsanalysen zur Vorhersage der Mathematikkompetenz (Klasse 10) durch fluide Fähigkeiten, Mathematikkompetenz und analytische Problemlösekompetenz (jeweils Klasse 9): Standardisierte Regressionskoeffizienten ( $\beta$ ; 95%-Konfidenzintervalle [KI] in eckigen Klammern) und aufgeklärte Varianz( $R^2$ ).....	117
<i>Tabelle 4.2</i>	Regressionsanalysen zur Vorhersage der naturwissenschaftlichen Kompetenz (Klasse 10) durch fluide Fähigkeiten, naturwissenschaftliche Kompetenz und analytische Problemlösekompetenz (jeweils Klasse 9): Standardisierte Regressionskoeffizienten ( $\beta$ ; 95%-Konfidenzintervalle [KI] in eckigen Klammern) und aufgeklärte Varianz ( $R^2$ ).....	120
<i>Tabelle 5.1</i>	Mittelwerte und Standardabweichungen (in Klammern) der in den Mehrebenenanalysen in Studie 4a verwendeten Variablen.....	145
<i>Tabelle 5.2</i>	Ergebnisse der Mehrebenenanalysen zur Vorhersage der Mathematikleistung (Studie 4a): Unstandardisierte Regressionskoeffizienten (95%-Konfidenzintervall in eckigen Klammern) und aufgeklärte Varianz.....	147
<i>Tabelle 5.3</i>	Mittelwerte und Standardabweichungen (in Klammern) der in der in den Mehrebenenanalysen in Studie 4b verwendeten Variablen.....	149
<i>Tabelle 5.4</i>	Ergebnisse der Mehrebenenanalysen zur Vorhersage der analytischen Problemlöseleistung (Studie 4b): Unstandardisierte Regressionskoeffizienten (95%-Konfidenzintervall in eckigen Klammern) und aufgeklärte Varianz.....	151

## Abbildungsverzeichnis

<i>Abbildung 1.1</i>	Testaufgabe „Anschlusszüge“ als Beispiel für Problemstellungen der Art Entscheidungen treffen.....	23
<i>Abbildung 1.2</i>	Testaufgabe „Ferienlager“ als Beispiel für Problemstellungen der Art Systeme analysieren und entwerfen.....	24
<i>Abbildung 1.3</i>	Testaufgabe „Bewässerung“ als Beispiel für Problemstellungen der Art Fehler suchen.....	25
<i>Abbildung 2.1</i>	Eindimensionales und dreidimensionales Modell der analytischen Problemlösekompetenz.....	57
<i>Abbildung 3.1</i>	Varianzen, Faktorladungen und Korrelationen für das Nested-Faktor-Modell, das erweiterte Standardmodell und das Einfaktormodell (standardisierte Koeffizienten; für alle Koeffizienten gilt: $p < .001$ ).....	89
<i>Abbildung 5.1</i>	Mittelwerte der Mathematikleistung (z-standardisiert) nach mathematischem Selbstkonzept (MSK; dichotomisiert am Median) und Kontextbedingung (Problemlösekontext [PL] versus Mathematikkontext [M]); $N = 256$ .....	148
<i>Abbildung 5.2</i>	Mittelwerte der Mathematikleistung (z-standardisiert) nach Mathematikangst (MA; dichotomisiert am Median) und Kontextbedingung (Problemlösekontext [PL] versus Mathematikkontext [M]); $N = 256$ .....	148



## Zusammenfassung

Die Ergebnisse der PISA-Studie 2003 zeigen für Schülerinnen und Schüler in Deutschland geringere Kompetenzen in der Mathematik und den Naturwissenschaften, als aufgrund ihrer (fächerübergreifenden) analytischen Problemlösekompetenz zu erwarten wären. Diese Diskrepanz kann als mangelnde Ausschöpfung des beim analytischen Problemlösen offensichtlich werdenden kognitiven Potenzials zum Aufbau fachbezogener Kompetenzen interpretiert werden (*Potenzialausschöpfungshypothese*). Um dieses Potenzial zur Förderung fachbezogener Kompetenzen didaktisch nutzen zu können, müssen zuvor zentrale Aspekte der Validität der Modellierung der analytischen Problemlösekompetenz eingehender untersucht werden. In den ersten drei Studien der Arbeit werden daher Aspekte der faktoriellen, diskriminanten, prognostischen und inkrementellen Validität der Modellierung der analytischen Problemlösekompetenz untersucht. In der vierten Studie wird mit der *Potenzialnutzungshypothese* eine ergänzende Erklärung für die bei PISA gefundenen Ergebnisse geprüft, sowie die Bedeutung motivationaler und emotionaler Faktoren für erfolgreiches analytisches Problemlösen untersucht. Die Ergebnisse deuten auf eine dreidimensionale Struktur der analytischen Problemlösekompetenz hin, und liefern zudem weitere Belege für die empirische Trennbarkeit der analytischen Problemlösekompetenz von fluiden Fähigkeiten, sowohl querschnittlich als auch im Längsschnitt. Sie zeigen ferner, dass analytische Problemlösekompetenz geeignet ist, zukünftige Kompetenzen in der Mathematik und den Naturwissenschaften auch über den Effekt der fluiden Fähigkeiten hinaus vorherzusagen. Zusätzlich konnte im Sinne der Potenzialnutzungshypothese gezeigt werden, dass eine mathematische Kontexteinbettung von sowohl analytischen Problemlöse- als auch von Mathematikaufgaben einen negativen Effekt auf die Leistungen in diesen Aufgaben haben kann. Dies gilt insbesondere für Schülerinnen und Schüler mit ungünstigen motivationalen und emotionalen Merkmalsausprägungen. Die Ergebnisse der Studien werden im Hinblick auf theoretische und praktische Implikationen sowie im Hinblick auf Limitationen und sich daraus ergebende Implikationen für zukünftige Forschung kritisch diskutiert.

## Abstract

Results from the PISA study 2003 show lower mean performances in mathematics and science for students in Germany than could be expected compared to their mean performance in (cross-curricular) analytical problem-solving. This discrepancy can be interpreted to the effect that students have a cognitive potential, which may not be fully exploited for the development of subject-related competencies (*cognitive potential exploitation hypothesis*). In order to make use of this potential to foster subject-related competencies, key aspects of the validity of the assessment of analytical problem solving competence have to be studied more profoundly. In the first three studies of this work, aspects of factorial, discriminant, prognostic, and incremental validity of the assessment of analytical problem solving are examined. In the fourth study, an additional explanation of the PISA results is introduced, the *cognitive potential application hypothesis*, and the significance of motivational and emotional factors for successful analytical problem solving are explored. The results point to a three-dimensional structure of analytical problem solving competence, and provide further evidence of the empirical distinction of analytical problem solving competence and fluid intelligence, in cross-sectional as well as in longitudinal perspective. Furthermore, the results show that analytical problem solving competence can predict future competencies in mathematics and science over and above fluid intelligence. In accordance with the cognitive potential application hypothesis, it was also demonstrated that a mathematical context embedding of analytical problem solving- as well as mathematics-tasks can have a negative effect on students' performance in those tasks. This is especially pronounced for students with unfavorable motivational and emotional dispositions. The results of all four studies are critically discussed with regard to theoretical and practical implications as well as limitations and consequences for future research.

# 1 ALLGEMEINE EINFÜHRUNG UND STRUKTUR DER ARBEIT

Die Fähigkeit, Probleme zu lösen, stellt eine zentrale Qualifikation in nahezu allen schulischen und außerschulischen Lern- und Leistungsbereichen dar. So wird Problemlösen als *domänen- oder fächerübergreifende* Kompetenz betrachtet, bei der die zur Lösung notwendigen Ressourcen nicht einer einzelnen Wissensdomäne zugeordnet werden können und die unter anderem für den beruflichen Erfolg eine zunehmend wichtigere Rolle spielt (Autor, Levy & Murnane, 2005; Binkley et al., 2012; OECD, 2004, 2014). Dementsprechend wird fächerübergreifende Problemlösekompetenz als ein zentrales Ziel schulischer Bildung angesehen (z. B. Binkley et al., 2012; Mayer, 2010; OECD, 2004) und ist folglich auch zum Testgegenstand von international vergleichenden Schulleistungstudien wie der PISA-Studie (*Programme for International Student Assessment*) der OECD (*Organisation for Economic Co-operation and Development*) geworden (z. B. Klieme, Leutner & Wirth, 2005; OECD, 2004, 2014).

Neben dieser domänen- oder fächerübergreifenden Perspektive (z. B. Bransford & Stein, 1984; Newell & Simon, 1972) wird Problemlösen auch als *fachbezogene* Kompetenz vor allem in der Mathematik, aber beispielsweise auch in den Naturwissenschaften betrachtet (Gabel & Bunce, 1994; Maloney, 1994; Reiss & Törner, 2007; Schoenfeld, 1985). Folglich wird Problemlösen in den Kerncurricula der Mathematik und der naturwissenschaftlichen Fächer Biologie, Chemie und Physik in Deutschland als zu unterrichtende beziehungsweise zu erwerbende fachbezogene Kompetenz definiert (z. B. Melle, Parchmann & Sumfleth, 2004; MSJK, 2004) und auch in den nationalen Bildungsstandards thematisiert (z. B. Blum, Drücke-Noe, Hartung & Köller, 2006; KMK, 2005a, 2005b, 2005c; für die USA siehe AAAS, 1993; NCTM, 1989; 2000).

## 1.1 Probleme und Problemlösen

Problemlösen hat als Forschungsgegenstand in der Psychologie, insbesondere in der Allgemeinen und Kognitiven Psychologie, eine lange, fast einhundertjährige Tradition. Die

Ursprünge liegen in den experimentellen Arbeiten der Gestalt- und Denkpsychologie der ersten Hälfte des 20. Jahrhunderts (z. B. Duncker, 1935; Wertheimer, 1945; für einen Überblick siehe Funke, 2003 sowie Mayer, 1992). Trotz unterschiedlicher Konzeptualisierungen der Begriffe *Problem* und *Problemlösen* herrscht weitgehende Einigkeit darüber, dass ein Problem aus einer Problemsituation (Anfangszustand), einem mehr oder weniger genau definierten Zielzustand und einem nicht unmittelbar einsichtigen Lösungsweg besteht (Hussy, 1998; Mayer, 1992). Dörner (1976) spricht in diesem Zusammenhang von einer Barriere zwischen Anfangs- und Zielzustand, sodass der Anfangszustand nicht ohne Weiteres in den Zielzustand überführt werden kann. Probleme stellen folglich, in Abgrenzung zu Aufgaben, keine Routine-Situationen dar, für die eine Lösung aus dem Gedächtnis abgerufen werden kann<sup>1</sup> (Funke, 2003). Das Lösen von Problemen lässt sich somit definieren als „zielorientiertes Denken und Handeln in Situationen, für deren Bewältigung keine routinierten Vorgehensweisen verfügbar sind“ (Klieme, Funke, Leutner, Reimann & Wirth 2001, S. 185; vgl. Mayer & Wittrock, 1996). Es geht, anders ausgedrückt, beim Problemlösen also darum, „Lücken in einem Handlungsplan zu füllen“ (Funke, 2003, S. 25). Nach Mayer (1990) zeichnet sich Problemlösen durch vier Charakteristika aus. Es ist (1) *personenbezogen*, da es von individuell verfügbaren Wissensinhalten, Strategien und Fähigkeiten abhängt (siehe die Ausführungen in Abschnitt 1.3.2), (2) *kognitiv*, da zur Lösung eines Problems eine kognitive Verarbeitung von Informationen notwendig ist, (3) *prozessbasiert*, da es die Anwendung von Wissen sowie das Erstellen und Verändern mentaler Repräsentationen erfordert (siehe die Ausführungen im folgenden Abschnitt 1.2) und (4) *gerichtet*, da es um das Erreichen eines angestrebten Zielzustandes geht. Sowohl in der Mathematik als auch in den Naturwissenschaften werden die Begriffe Problem und Problemlösen in ähnlicher Art und Weise verwendet und primär auf Grundlage der oben aufgeführten psychologischen Definitionen rezipiert (z. B. Bodner & Herron, 2002; Maloney, 1994; Schoenfeld, 1985, 1992).

---

<sup>1</sup> Wenn im Folgenden von Problemlöseaufgaben die Rede ist, so sind damit nicht Aufgaben im Sinne von Routinesituationen gemeint, für die eine Lösung aus dem Gedächtnis abgerufen werden kann (vgl. Funke, 2003), sondern tatsächlich Problemstellungen, für welche die hier genannten Kriterien gelten. Der Einfachheit halber und dem üblichen Sprachgebrauch folgend sollen jedoch im Kontext von Testinstrumenten zur Erfassung des Problemlösens die Begriffe *Aufgabe* und *Problemstellung* synonym verwendet werden.

Zur Klassifizierung von Problemen wurden, insbesondere in der Kognitionspsychologie, verschiedene Taxonomien erstellt (siehe Frensch & Funke, 1995). McCarthy (1956) unterscheidet beispielsweise zwischen gut definierten (*well-defined*) und schlecht definierten (*ill-defined*) Problemen (siehe auch Reitman, 1965; Sternberg & Davidson, 1992). Jonassen (2000) erweitert diese Klassifikation und unterscheidet elf Problemklassen auf einem Kontinuum zwischen *gut strukturiert* und *schlecht strukturiert*. Arbing (1997) schlägt eine Differenzierung je nach benötigtem Wissen vor. Er unterscheidet Probleme, die zur Lösung lediglich heuristische Methoden erfordern, von solchen Problemen, die eine enge Verknüpfung von Heuristiken und (fach-)spezifischem Wissen erfordern (vgl. Robertson, 2001). In ähnlicher Weise unterscheiden Chi, Glaser und Rees (1982) *semantisch reiche* Probleme, die zur Lösung umfangreiches Vorwissen erfordern, von *semantisch armen* Problemen, die kein oder kaum Vorwissen erfordern. Dörner klassifiziert Problemtypen nach der Art der Barriere, die beim Lösen des Problems überwunden werden muss. Diese hängt vom „Bekanntheitsgrad der Mittel“ und der „Klarheit der Zielkriterien“ ab (Dörner, 1976, S. 14; vgl. auch Arlin, 1989 sowie Wagener, 2001). Im Kontext der empirischen Bildungsforschung sowie der pädagogischen Praxis in Schule und Unterricht stehen vor allem gut strukturierte Probleme mit klaren Zielkriterien im Fokus, für deren Lösung mehr oder weniger spezifisches Vorwissen notwendig ist (z. B. Lorenzo, 2005; OECD, 2004, 2014; Reiss & Törner, 2007). Diese Art von Problemstellungen ist auch Gegenstand der empirischen Studien der vorliegenden Arbeit.

Insbesondere im Kontext der empirischen Forschung zum Problemlösen im Rahmen von Large-Scale-Assessments wird im Hinblick auf die Art einer Problemstellung zwischen *analytischen* und *dynamischen* Aspekten des Problemlösens unterschieden (Klieme et al., 2001; Leutner, Klieme, Meyer & Wirth, 2004; Wirth & Klieme, 2003). Das *analytische* Problemlösen zeichnet sich dadurch aus, dass alle für die Problemlösung relevanten Informationen in der Problemstellung explizit gegeben sind oder erschlossen werden können. Problemlösen kann damit zu einem großen Teil als schlussfolgernde Anwendung vorhandenen Wissens bezeichnet werden (Leutner, Funke, Klieme & Wirth, 2005). Beim *dynamischen* Problemlösen hingegen muss ein Großteil der lösungsrelevanten Informationen in einer explorierenden Interaktion mit der Problemsituation erst generiert werden (*learning*

*by doing*; Wirth & Klieme, 2003; vgl. Greiff, Wüstenberg & Funke, 2012). Derartige dynamische Problemstellungen werden in der kognitionspsychologischen Forschung häufig unter dem Begriff *komplexe Probleme* behandelt (Funke, 2003; Funke & Frensch, 2007) und im Kontext von PISA 2012 als *interaktive Problemstellungen* bezeichnet (Greiff & Funke, 2017; OECD, 2014). Demgegenüber werden analytische Problemstellungen in der kognitionspsychologischen Forschung mitunter als *einfache Probleme* (Hussy, 1998; Knoblich & Öllinger, 2006) und in PISA 2012 als *statische Problemstellungen* bezeichnet (OECD, 2014; Ramalingam, Philpot & McCrae, 2017). Beide Arten von Problemstellungen, interaktive und statische, werden wiederum als Komponenten des *kreativen Problemlösens* verstanden (OECD, 2013, 2014), welches häufig als eines der zentralen Bildungsziele im 21. Jahrhundert proklamiert wird (Binkley et al., 2013; OECD, 2014). Die Unterscheidung von dynamischen (komplexen oder interaktiven) und analytischen (einfachen oder statischen) Aspekten des Problemlösens bezieht sich ausschließlich auf die genannten Merkmale der entsprechenden Problemstellungen und impliziert keine Unterschiede in den Lösungsschwierigkeiten dieser Problemstellungen. Es kann also nicht davon ausgegangen werden, dass komplexe Probleme per se schwieriger zu lösen sind als analytische Probleme oder umgekehrt. Gegenstand der empirischen Studien der vorliegenden Arbeit sind fächerübergreifende analytische Problemstellungen, wie sie im Rahmen von PISA 2003 eingesetzt wurden (siehe Abschnitt 1.3.3 für eine Beschreibung des Problemlösetests in PISA 2003).

## 1.2 Problemlösen als Prozess

Basierend auf den Überlegungen der Gestaltpsychologie (z. B. Duncker, 1935) und den Arbeiten von Dewey (1910) beschreibt der ungarische Mathematiker Polya (1945) in seiner klassischen Arbeit Problemlösen als eine Abfolge von vier Prozessschritten: (1) *understanding the problem*, (2) *devising a plan*, (3) *carrying out the plan*, und (4) *looking back* (i. S. v. Bewertung und Diskussion der Problemlösung). Polya liefert damit die Grundlage für zahlreiche jüngere Konzeptualisierungen des Problemlösens sowohl im Sinne des domänen- und fächerübergreifenden Problemlösens in der Psychologie als auch im Sinne des fachbezogenen Problemlösens vor allem in der Mathematik, aber auch in den Naturwissenschaften. So unterscheidet Mayer (1992) in seinem Modell des Problemlösens zwischen

unterschiedlichen Wissensarten und den kognitiven Prozessen *Problemrepräsentation* und *Problemlösen*, welche für eine erfolgreiche Problemlösung notwendig sind. Der Prozess der Problemrepräsentation lässt sich weiter differenzieren in das Übersetzen eines Problems in eine mentale Repräsentation, welche in eine kohärente mentale Struktur überführt wird. Der Prozess des Problemlösens unterteilt sich ebenfalls in zwei Teilprozesse: das Planen des Lösungsweges und das Überwachen der Durchführung.

Mit Bezug auf Polya unterscheidet Schoenfeld (1983) die Prozessschritte *reading*, *analysis*, *exploration*, *planning*, *implementation* und *verification* des Lösens mathematischer Probleme. Carlson und Bloom (2005) entwickeln ebenfalls, aufbauend auf den Überlegungen von Polya, ein Modell des mathematischen Problemlösens und unterscheiden dabei die Prozesse *orientation*, *planning*, *executing* und *checking*. Mathematisches Problemlösen lässt sich somit in ähnlicher Art und Weise wie fächerübergreifendes Problemlösen als eine Abfolge von Prozessschritten beschreiben (siehe hierzu auch Artzt & Armour-Thomas, 1992). Eine vergleichbare Abfolge von Prozessschritten findet sich auch in der Beschreibung des mathematischen Modellierens. Der Modellierungskreislauf, der vor allem von Freudenthal (1983) und Schupp (1988) beschrieben wurde und der die theoretische Basis für die Mathematiktests in den PISA-Studien darstellt (Klieme, Neubrand & Lüdtke, 2001; OECD, 2003, 2013), unterscheidet sich kaum von Modellen des mathematischen Problemlösens (z. B. Carlson & Bloom, 2005; Lester & Kehle, 2003). Das Schema des mathematischen Modellierens geht von einer Problemsituation in der realen Welt aus, welche verstanden und schließlich „mathematisiert“, das heißt in Begriffen der Mathematik beschrieben werden muss (Schupp, 1988). Im Anschluss müssen mathematische Hilfsmittel ausgewählt und angewendet werden. Das mathematische Ergebnis muss schließlich interpretiert und an der Ausgangssituation validiert werden (für Erweiterungen des klassischen Modellierungskreislaufs siehe z. B. Blomhoj & Hojgaard Jensen, 2003; Staub & Reusser, 1995). Somit kann mathematisches Modellieren als mathematisches Problemlösen im weiteren Sinne beschrieben werden, während mathematisches Problemlösen im engeren Sinne einen rein innermathematischen Prozess bezeichnet. Sofern jedoch keine Unterscheidung zwischen inner- und außermathematischer Problemsituation getroffen wird, sind beide Konzeptionen praktisch deckungsgleich. Somit lassen sich die beim mathemati-

schen Problemlösen ablaufenden Prozesse je nach gewählter Fokussierung mit unterschiedlichen Begrifflichkeiten beschreiben (Blum & Niss, 1991; Leiß & Blum, 2006; Neubrand et al., 2001).

Wie für die Mathematik beschrieben stellt auch für die Naturwissenschaften die Arbeit von Polya (1945) eine grundlegende Struktur für die Einteilung des Problemlösens in Phasen dar (z. B. Gabel & Bunce, 1994; Bunce & Heikkinen, 1986; Maloney, 1994). So übertragen Mettes, Pilot, Roossink und Kramers-Pals (1980) die Phasen des Problemlösens nach Polya auf den Bereich der sogenannten „specification problems in science“ und unterscheiden die vier Phasen: (1) *analysis of the problem*, (2) *planning the problem solving process*, (3) *execution of routine operations* und (4) *checking the answer and interpretation of the results*. Lorenzo (2005) entwickelt ebenfalls aufbauend auf den Prozessschritten nach Polya ein Modell des Problemlösens in der Chemie, welches zwölf Prozessschritte umfasst. Darüber hinaus zeigen die beschriebenen Prozessschritte des naturwissenschaftlichen Problemlösens eine hohe Kongruenz zu den Phasen des wissenschaftlichen Denkens (*scientific reasoning*; z. B. Dunbar, 1995) und der naturwissenschaftlichen Erkenntnisgewinnung (*scientific inquiry*; z. B. Bybee, 2000). Für das Kerncurriculum Biologie der gymnasialen Oberstufe in Deutschland (Harms, Mayer, Hammann, Bayrhuber & Kattmann, 2004) lassen sich exemplarisch vier Phasen des wissenschaftlichen Denkens unterscheiden: (1) Fragestellung und Hypothesenbildung (i. S. v. *analysis of the problem*), (2) Untersuchungsplanung (i. S. v. *planning the problem solving process*), (3) Untersuchungsdurchführung (i. S. v. *carrying out a plan*) und (4) Auswertung und Deutung der Befunde (i. S. v. *checking the answer and interpretation of the results*). Problemlösen kann demnach als die paradigmatische naturwissenschaftliche Erkenntnismethode angesehen werden (Dunbar & Fugelsang, 2005; Mayer 2007).

Es lässt sich festhalten, dass die kognitiven Prozesse, die zum Verstehen einer Problemsituation und zur erfolgreichen Bearbeitung eines Problems notwendig sind, als eine Abfolge von Prozessschritten beschrieben werden können (z. B. Bransford & Stein, 1984; Dewey, 1910; Hayes, 1981; Polya, 1945; Sternberg & Davidson, 1992). Allgemein muss ein Problem zunächst als solches identifiziert und verstanden werden, das heißt es müssen relevante Informationen und Bedingungen des Problems erkannt und diese in eine mentale Repräsentation übersetzt werden (z. B. Kintsch & Greeno, 1985; Staub & Reusser, 1995).



Im Anschluss daran erfolgt die Planung eines Lösungsweges und dessen Anwendung. Schließlich ist die Lösung des Problems zu überprüfen und das Ergebnis zu kommunizieren (vgl. OECD, 2004, Kap. 2).

## 1.3 Problemlösen als Kompetenz

### 1.3.1 Der Kompetenzbegriff

Die Vermittlung von Wissen und Fertigkeiten in konkreten Anforderungsbereichen, die häufig unter dem Begriff *Kompetenz* gefasst werden, ist ein zentrales Anliegen schulischer und beruflicher Bildung (Klieme, Hartig & Rauch, 2008; Köller 2009). Der Messung von Kompetenzen kommt damit eine Schlüsselfunktion für die Optimierung von Bildungsprozessen sowie für die Qualitätssicherung und Qualitätsentwicklung des Bildungswesens zu (Fleischer, Koeppen, Kenk, Klieme & Leutner, 2013; Klieme & Leutner 2006; Koeppen, Hartig, Klieme & Leutner, 2008). Die vielfältige Bedeutung des Kompetenzbegriffs, der – obwohl keineswegs neu – erst seit Beginn der 2000er Jahre zum Gegenstand intensiver Diskussionen in der Psychologie und der Bildungsforschung geworden ist (z. B. Csapó 2004; Hartig & Klieme 2006; Klieme et al., 2008), hat Weinert (2001) systematisch aufbereitet. Er unterscheidet dabei sechs Definitionen des Begriffs. So werden Kompetenzen unter anderem als „auf spezifische Kontexte bezogene kognitive Leistungsdispositionen“, aber beispielsweise auch als „für die Bewältigung anspruchsvoller Aufgaben notwendige motivationale Orientierungen“ definiert (vgl. auch Klieme 2004). Aufgrund theoretischer und pragmatischer Überlegungen empfiehlt Weinert (2001) eine Einschränkung auf erstgenannte Definition. Dieser Empfehlung folgend definieren Klieme und Leutner (2006) im Rahmen des DFG-Schwerpunktprogramms „Kompetenzmodelle“ Kompetenzen als „kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen“ (Klieme & Leutner 2006, S. 879) und die durch Erfahrung und Lernen sowie äußere Interventionen und institutionalisierte Bildungsprozesse erworben und beeinflusst werden (Klieme & Leutner 2006; Klieme et al., 2008). Mit dieser Definition sind zwei Einschränkungen verbunden: Zum einen sind Kompetenzen bereichsspezifisch auf bestimmte Kontexte beziehungsweise Anforderungssituationen bezogen, wodurch allgemeine kognitive Fähigkeiten außen vor bleiben. Kompetenzen werden damit,

in einem größeren Maße als dies für allgemeine kognitive Fähigkeiten der Fall ist, als prinzipiell erlernbar und trainierbar verstanden. Zum anderen erfolgt eine Beschränkung auf kognitive Aspekte, wodurch motivationale und emotionale Faktoren ebenfalls zunächst ausgeschlossen werden. Diese Einschränkungen erfolgen vor allem aus pragmatischen Erwägungen und sollen nicht die prinzipielle Bedeutung allgemeiner kognitiver Grundfähigkeiten wie zum Beispiel der Intelligenz oder motivationaler Einstellungen und Orientierungen sowie emotionaler Faktoren für erfolgreiches Handeln in spezifischen Anforderungssituationen in Abrede stellen<sup>2</sup> (Fleischer et al., 2013; Wilhelm & Nickolaus, 2013). Die Definition des Kompetenzbegriffs nach Klieme und Leutner (2006) steht in wesentlichen Punkten in Übereinstimmung mit der Verwendung des Kompetenzbegriffs im Kontext internationaler Schulleistungsstudien (z. B. PISA, TIMSS, IGLU/PIRLS) sowie der KMK-Bildungsstandards (z. B. Pant, Böhme & Köller, 2012). So werden Kompetenzen in den PISA-Studien als „prinzipiell erlernbare, mehr oder minder bereichsspezifische Kenntnisse, Fertigkeiten und Strategien“ definiert (Baumert, Stanat & Demmrich, 2001, S. 22).

### 1.3.2 Komponenten der Problemlösekompetenz

Im Verlauf der im Abschnitt 1.2 skizzierten Prozessschritte des Problemlösens lassen sich verschiedene kognitive Komponenten identifizieren, die für eine erfolgreiche Bearbeitung sowohl fächerübergreifender analytischer als auch fachbezogener Problemstellungen relevant sind (O’Neil, 1999; Sugrue, 1995). Hierzu gehört zunächst die Verfügbarkeit und Anwendung von *inhaltsspezifischem Wissen* (Schoenfeld, 1985; Süß, 1996). Nach Süß (1996) lässt sich dabei Sach- von Handlungswissen unterscheiden, welches jeweils sowohl deklarativ als auch prozedural repräsentiert sein kann. *Sachwissen* ist definiert als Wissen über Objekte und Zustände, mit dessen Hilfe die Problemsituation und der angestrebte Zielzustand intern repräsentiert und damit erfasst werden können (*knowing that*; Resnick, 1983). *Handlungswissen* ist definiert als Wissen über Operationen zur Veränderung der Problemsituation und die Überführung dieser in den angestrebten Zielzustand sowie die Fähigkeit, eine kognitive Operation oder Handlung auch tatsächlich auszuführen (*knowing*

---

<sup>2</sup> Studie 4 der vorliegenden Arbeit (siehe Kapitel 5) greift diesen Punkt auf, indem die Bedeutung motivationaler und emotionaler Faktoren für die analytische Problemlösekompetenz in zwei experimentellen Studien untersucht wird.

*how*; Resnick, 1983; vgl. Anderson, 1982). Paris, Lipson und Wixson (1983) ergänzen *konditionales Wissen* (*knowing when and why*; Paris, Lipson & Wixson, 1983), welches die Umstände der Anwendung von Operationen beschreibt (für eine Darstellung unterschiedlicher taxonomischer Ordnungen des Wissensbegriffs sei außerdem auf Dochy & Alexander, 1995, sowie Rittle-Johnson, Siegler & Alibali, 2001, verwiesen).

Neben inhaltsspezifischem Sach-, Handlungs- und konditionalem Wissen sind die Verfügbarkeit und die Anwendung von *allgemeinen Problemlösestrategien* und *Heuristiken*, welche die Suche nach relevanten Informationen, alternativen Problemrepräsentationen, Analogien oder Teilzielen strukturieren, eine weitere Komponente erfolgreichen Problemlösens (Gick, 1986). Strategien und Heuristiken wurden insbesondere im Rahmen des Informationsverarbeitungsansatzes des Problemlösens untersucht (z. B. Atwood & Polson, 1976; Dörner, 1976). In ihrer klassischen Arbeit beschreiben Newell und Simon (1972) Problemlösen als die Suche in einem Problemraum, welcher alle möglichen Zustände enthält, die bei der Lösung eines Problems auftreten können. Newell und Simon gehen davon aus, dass Menschen diesen Problemraum mithilfe allgemeiner Heuristiken gezielt durchsuchen, um zur Problemlösung zu gelangen. Zwei der wichtigsten Heuristiken sind dabei die Methode der Unterschiedsreduktion (*hill climbing*), bei der immer derjenige mentale Operator eingesetzt wird, der den aktuellen Zustand dem angestrebten Zielzustand am nächsten bringt, sowie die Mittel-Ziel-Analyse (*means-end analysis*), bei der Teilziele zwischen Ist- und Sollzustand definiert werden, welche dann durch Anwendung geeigneter mentaler Operatoren nach und nach abgearbeitet werden (für Erweiterungen des Problemraummodells und seine Anwendung auf naturwissenschaftliches Arbeiten siehe z. B. Klahr & Dunbar, 1988).

Neben der Verfügbarkeit und Anwendung von Wissensinhalten und Problemlösestrategien stellt die Fähigkeit zur *Selbstregulation* eine weitere Komponente der Problemlösekompetenz dar, durch die Prozesse des Problemlösens geplant, überwacht, bewertet und gegebenenfalls modifiziert werden (vgl. Berardi-Coletta, Buyer, Dominowski & Rellinger, 1995; Davidson, Deuser & Sternberg, 1994; Davis, 2000). An dieser Stelle sei darauf hingewiesen, dass die Begriffe Selbstregulation und Metakognition in der Literatur mit unterschiedlicher, aber teilweise überlappender Bedeutung verwendet werden, ebenso wie das

weiter oben beschriebene Konzept des konditionalen Wissens auch mit den Begrifflichkeiten der Metakognitionsforschung beschrieben werden kann (z. B. Schneider, Schlagmüller & Visé, 1998). Entscheidend für eine erfolgreiche Selbstregulation des Problemlöseprozesses ist die Fähigkeit, das eigene Vorgehen zu beobachten und zu bewerten (Monitoring-Komponente) sowie gegebenenfalls zu regulieren (Regulations-Komponente; siehe Schreiber, 1998). Hierbei spielen auch selbstbezogene Kognitionen wie Fähigkeitsselbstkonzepte (z. B. Marsh, 1992), wie sie in Studie 4 der vorliegenden Arbeit untersucht werden (siehe Kapitel 5), oder Selbstwirksamkeitserwartungen eine Rolle (z. B. Zimmerman, Bandura & Martinez-Pons, 1992). Beispielsweise ist anzunehmen, dass sich Personen mit einem geringen Selbstkonzept und geringen Selbstwirksamkeitserwartungen weniger anspruchsvolle Teilziele setzen oder die Lösung eines Problems erst gar nicht in Angriff nehmen (Bandura, 1997; Bouffard-Bouchard, Parent & Larivée, 1991).

Um die Bedeutung dieser Komponenten für fächerübergreifende analytische und fachbezogene Problemlösekompetenz zu untersuchen, analysierten Fleischer, Wirth, Rumann und Leutner (2010) die kognitiven Anforderungen der entsprechenden PISA-2003-Testaufgaben zur Erfassung der analytischen Problemlösekompetenz und der Mathematikkompetenz. Ziel der Analyse war es, möglichst umfassend konzeptuelle und strukturelle Gemeinsamkeiten und Unterschiede dieser Testaufgaben zu identifizieren (für einen entsprechenden Vergleich der Problemlöseaufgaben mit den Testaufgaben zur Erfassung der naturwissenschaftlichen Kompetenz in PISA 2003 siehe Rumann, Fleischer, Stawitz, Wirth & Leutner, 2010). Die Ergebnisse zeigen unter anderem, dass sich die Problemlöseaufgaben durch eine detailliertere Beschreibung der Ausgangssituation und einen stärkeren persönlichen und alltäglichen Kontextbezug auszeichnen als die Mathematikaufgaben. Obwohl die meisten Problemlöseaufgaben keiner fachlichen Domäne zugeordnet werden können, weist etwa ein Drittel der Aufgaben einen klaren Mathematikbezug auf. Die Mathematikaufgaben weisen einen stärkeren Formalisierungsgrad auf und stellen damit höhere Anforderungen an die Fähigkeit zur Dekodierung lösungsrelevanter Informationen. Hierfür ist inhaltspezifisches Sachwissen notwendig, was sich auch am höheren curricularen Niveau des zur Lösung der Mathematikaufgaben notwendigen Wissens zeigt. Bei den Problemlöseaufgaben hingegen scheint eher Handlungswissen eine relevante Komponente zu sein.

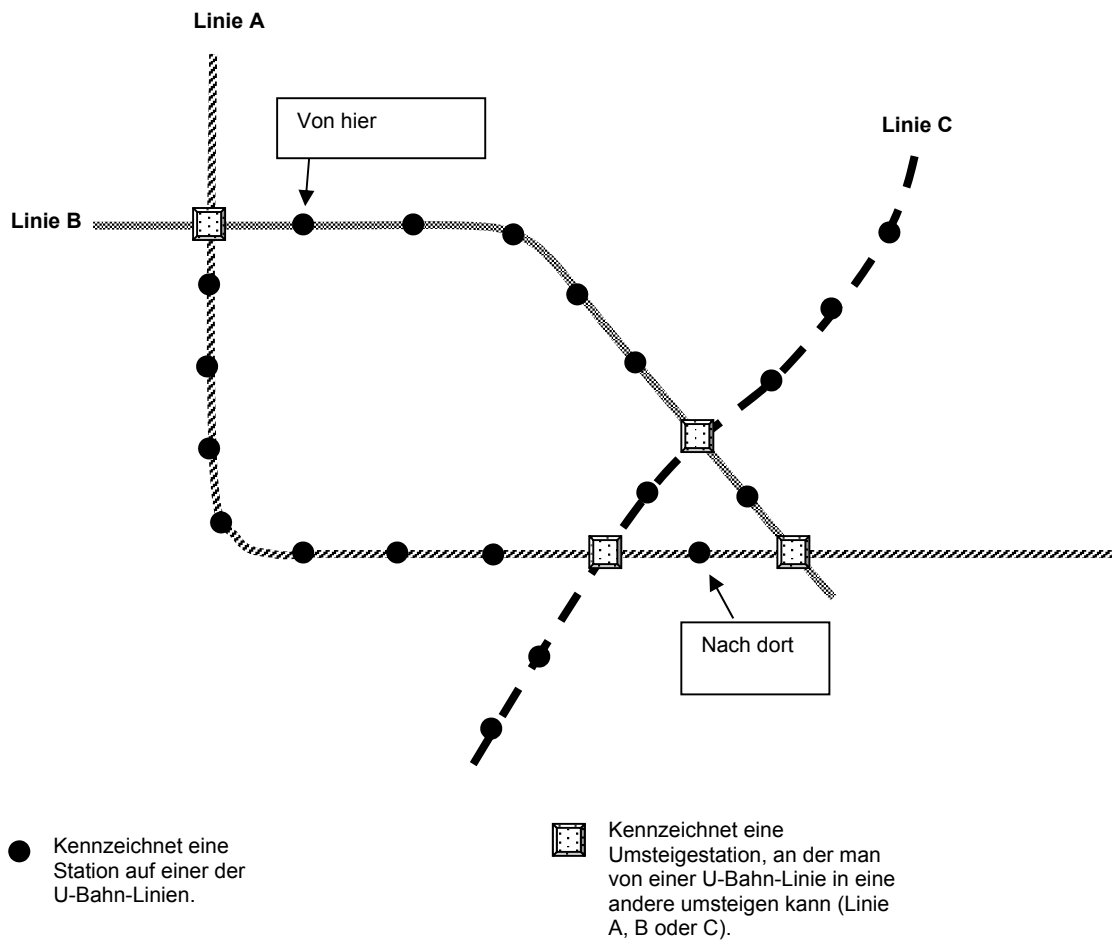
Insbesondere die höheren Anforderungen an das Erkennen lösungsrelevanter Bedingungen und an das planvolle und strategische Vorgehen deuten auf die Relevanz konditionalen Wissens und der Fähigkeit zur Selbstregulation hin. Dies scheinen eher relevante Komponenten der fächerübergreifenden und weniger der mathematischen Problemlösekompetenz zu sein (Fleischer et al., 2010).

### 1.3.3 Problemlösekompetenz und fachbezogene Kompetenzen in PISA 2003

Gegenstand der PISA-Erhebungen ist die Erfassung von Kompetenzen 15-jähriger Schülerinnen und Schüler im internationalen Vergleich. Fächerübergreifende Problemlösekompetenz von Schülerinnen und Schülern wurde in der PISA-Studie 2003 in Form der *fächerübergreifenden analytischen Problemlösekompetenz* erfasst. Dieser Test misst die „Fähigkeit einer Person, kognitive Prozesse zu nutzen, um sich mit solchen realen, fächerübergreifenden Problemstellungen auseinanderzusetzen und sie zu lösen, bei denen der Lösungsweg nicht unmittelbar erkennbar ist und die zur Lösung nutzbaren Wissensbereiche nicht einem einzelnen Fachgebiet der Mathematik, der Naturwissenschaften oder des Lesens entstammen“ (Leutner et al., 2004, S. 148; siehe auch OECD, 2003, S. 156). Der Test umfasst zehn Aufgaben beziehungsweise Problemstellungen mit teilweise mehreren Fragestellungen beziehungsweise Handlungsanweisungen pro Aufgabe, woraus 18 Testitems als Analyseeinheiten resultieren. Die Testitems gingen teilweise als dichotome (0 = *falsch*; 1 = *richtig*) und teilweise als polytome Items (0 = *falsch*; 1 = *teilweise richtig*; 2 = *vollständig richtig*) in eine eindimensionale Skalierung der Kompetenzskala ein. Das spezifizierte Testmodell entspricht damit dem Partial-Credit-Modell (Masters, 1982), einer Erweiterung des dichotomen Rasch-Modells (Rasch, 1960) für ordinale Daten (OECD, 2005). Auf theoretisch-konzeptueller Ebene lassen sich drei unterschiedliche Arten von verwendeten Problemstellungen differenzieren, die sich hinsichtlich der zu erreichenden Ziele, der beteiligten Prozesse und der potenziell schwierigkeiterzeugenden Merkmale unterscheiden können: (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* (Leutner et al., 2004, OECD, 2003, Kap. 4). Abbildung 1.1 stellt die Testaufgabe „Anschlusszüge“ als Beispiel für eine Problemstellung der Art *Entscheidungen treffen* dar, bei der es darum geht, die schnellste und kostengünstigste Verbindung zwischen zwei U-Bahn-

Stationen zu bestimmen. Problemstellungen der Art *Systeme analysieren und entwerfen* lassen sich am Beispiel der Testaufgabe „Ferienlager“ verdeutlichen, die in Abbildung 1.2 dargestellt ist. Bei dieser Aufgabe geht es darum, Jungen und Mädchen sowie Erwachsenen verschiedene Schlafsäle zuzuweisen, wobei einschränkende Randbedingungen hinsichtlich des zahlenmäßigen Verhältnisses Erwachsener zu Kind sowie der Geschlechterzusammensetzung und der Anzahl der Betten pro Schlafsaal berücksichtigt werden müssen. Abbildung 1.3 stellt die Testaufgabe „Bewässerung“ als Beispiel für eine Problemstellung der Art *Fehler suchen* dar. Diese Aufgabe besteht aus mehreren Testitems, bei denen es darum geht, zunächst die Funktionsweise eines Systems von Bewässerungskanälen, welches durch Schleusen gesteuert wird, zu verstehen und im Anschluss verschiedene Fehlfunktionen dieses Systems zu identifizieren (siehe auch OECD, 2004). Für eine eingehendere Darstellung der drei unterschiedlichen Arten von Problemstellungen in PISA 2003 sei auf die Studie 1 in Kapitel 2 verwiesen, in der untersucht wird, ob diese drei unterschiedlichen Arten von Problemstellungen empirisch voneinander trennbare Dimensionen abbilden.

Die folgende Abbildung zeigt einen Ausschnitt des öffentlichen Verkehrsnetzes einer Stadt in Zedland mit drei U-Bahn-Linien. Der Ort, an dem du dich zur Zeit befindest, sowie dein Zielort sind eingezeichnet.



Der Preis richtet sich nach der Anzahl der angefahrenen Stationen (die Abfahrtsstation nicht mitgerechnet). Die Kosten betragen 1 Zed pro angefahrter Station.

Die Fahrzeit zwischen zwei aufeinander folgenden Stationen beträgt ungefähr 2 Minuten. Um an einer Umsteigestation von einer U-Bahn-Linie in eine andere umzusteigen, benötigt man ungefähr 5 Minuten.

Die Abbildung zeigt die Station, an der du dich zur Zeit befindest („Von hier“), und die Station, zu der du fahren möchtest („Nach dort“). **Markiere in der Abbildung** die beste Strecke in Bezug auf Kosten und Zeit und nenne nachfolgend den Fahrpreis sowie die ungefähre Fahrzeit.

Fahrpreis: ..... Zeds.

Ungefähre Fahrzeit: ..... Minuten.

**Abbildung 1.1: Testaufgabe „Anschlusszüge“ als Beispiel für Problemstellungen der Art Entscheidungen treffen**

Die Gemeinde von Zedstadt organisiert ein fünftägiges Ferienlager für Kinder. 46 Kinder (26 Mädchen und 20 Jungen) haben sich für das Ferienlager angemeldet, und 8 Erwachsene (4 Männer und 4 Frauen) haben sich als Freiwillige gemeldet, um sie zu betreuen und die Freizeit zu organisieren.

**Tabelle 1: Erwachsene**

Frau Marjan
Frau Christen
Frau Grethen
Frau Karina
Herr Simon
Herr Niklas
Herr Wilhelm
Herr Peters

**Tabelle 2: Schlafsäle**

Name	Anzahl der Betten
Rot	12
Blau	8
Grün	8
Lila	8
Orange	8
Gelb	6
Weiß	6

1. Jungen und Mädchen müssen in getrennten Schlafsälen schlafen.
2. In jedem Schlafsaal muss mindestens ein Erwachsener schlafen.
3. Der/die Erwachsene/n in einem Schlafsaal muss/müssen das gleiche Geschlecht haben wie die Kinder.

Vervollständige die Tabelle zur Verteilung der 46 Kinder und der 8 Erwachsenen auf die Schlafsäle. Beachte alle Vorgaben.

Name	Anzahl der Jungen	Anzahl der Mädchen	Name/n der/des Erwachsenen
Rot			
Blau			
Grün			
Lila			
Orange			
Gelb			
Weiß			

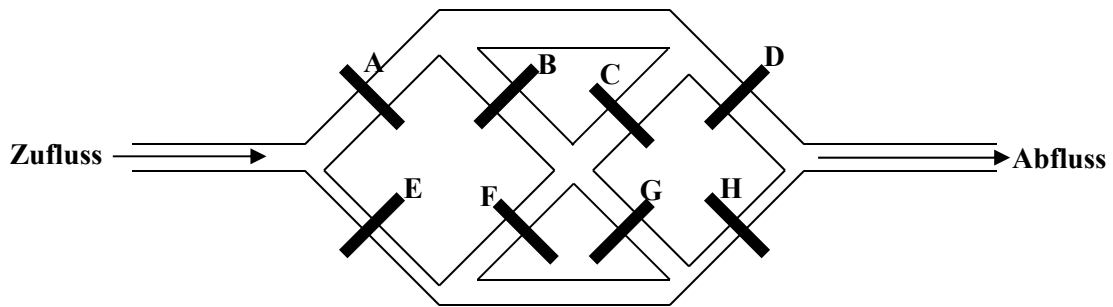
**Abbildung 1.2: Testaufgabe „Ferienlager“ als Beispiel für Problemstellungen der Art**  
**Systeme analysieren und entwerfen**



Die folgende Abbildung zeigt ein System von Bewässerungskanälen. Sie dienen zur Bewässerung für Getreideflächen. Die Schleusentore A bis H können geöffnet oder geschlossen werden, um das Wasser dorthin zu leiten, wo es gebraucht wird. Wenn ein Schleusentor geschlossen ist, kann kein Wasser durchfließen.

Bei dieser Aufgabe geht es darum, ein Schleusentor zu ermitteln, das geschlossen ist und klemmt und das Wasser daran hindert, durch das Kanalsystem zu fließen.

**Abbildung 1: Ein System von Bewässerungskanälen**



**Tabelle 1: Einstellung der Schleusentore**

A	B	C	D	E	F	G	H
Offen	Geschlossen	Offen	Offen	Geschlossen	Offen	Geschlossen	Offen

Michael stellt fest, dass bei Einstellung der Schleusentore nach Tabelle 1 kein Wasser durchfließt, was darauf hindeutet, dass mindestens eines der auf „Offen“ stehenden Schleusentore tatsächlich klemmt und geschlossen bleibt.

Beurteile für jeden der folgenden Problemfälle, ob das Wasser bis zum Ende fließt. Kreise für jeden Fall „Ja“ oder „Nein“ ein.

Problemfall	Fließt Wasser bis zum Ende?
Das Schleusentor <b>A</b> klemmt und bleibt geschlossen. Alle anderen Schleusentore funktionieren richtig wie in Tabelle 1 gesetzt.	Ja / Nein
Das Schleusentor <b>D</b> klemmt und bleibt geschlossen. Alle anderen Schleusentore funktionieren richtig wie in Tabelle 1 gesetzt.	Ja / Nein
Das Schleusentor <b>F</b> klemmt und bleibt geschlossen. Alle anderen Schleusentore funktionieren richtig wie in Tabelle 1 gesetzt.	Ja / Nein

**Abbildung 1.3: Testaufgabe „Bewässerung“ als Beispiel für Problemstellungen der Art Fehler suchen**

Der in PISA 2003 eingesetzte Test zur Erfassung der Mathematikkompetenz basiert auf dem Konzept der *Mathematical Literacy* und erfasst, „inwieweit Schülerinnen und Schüler [...] Mathematik in problemhaltigen außer- oder innermathematischen Situationen flexibel anwenden können“ (Blum et al., 2004, S. 47). Dieser Test weist deutliche theoretisch-konzeptuelle Ähnlichkeiten mit dem zuvor beschriebenen Test zur Erfassung der analytischen Problemlösekompetenz auf. Diese Ähnlichkeiten betreffen sowohl die zur erfolgreichen Bearbeitung von Problemlöseaufgaben und Mathematikaufgaben erforderlichen Prozessschritte als auch die im Ablauf des Lösungsprozesses relevanten Komponenten und Anforderungen (Fleischer et al., 2010; siehe hierzu auch die Ausführungen in Abschnitt 1.2 zu den Prozessschritten des Problemlösens sowie die Ausführungen in Abschnitt 1.3.2 zu den Komponenten des Problemlösens). So beschreibt das PISA 2003-Assessment-Framework für den Problemlösetest folgende Prozessschritte, die durchlaufen werden müssen, um eine Problemstellung erfolgreich zu bearbeiten: (1) *understand*, (2) *characterize*, (3) *represent and* (4) *solve the problem*, (5) *reflect and* (6) *communicate the problem solution* (OECD, 2003, S. 170 f.). Eine vergleichbare Darstellung von Prozessschritten findet sich auch bei der Beschreibung des mathematischen Modellierens (z. B. Freudenthal, 1983), welches die theoretische Basis für den PISA-Mathematiktest darstellt: (1) *starting with a problem situated in reality*, (2) *organizing it according to mathematical concepts*, (3) *transforming the real problem into a mathematical problem: generalizing and formalizing*, (4) *solving the mathematical problem and* (5) *making sense of the mathematical solution in terms of the real situation* (OECD, 2003, S. 27; vgl. Blum et al., 2004). Während die analytische Problemlösekompetenz nur geringes mathematisches Wissen erfordert, verlangt die Mathematikkompetenz erwartungsgemäß elaboriertere mathematische Kenntnisse und das Beherrschen auch komplexerer mathematischer Prozeduren. Darüber hinaus werden analytische Problemlösekompetenz und Mathematikkompetenz als relativ unabhängig von Lesekompetenz und naturwissenschaftlichem Wissen, jedoch mit deutlichem Bezug zu schlussfolgerndem Denken konzeptualisiert (OECD, 2003, 2004). Die theoretisch-konzeptuellen Ähnlichkeiten zwischen analytischer Problemlösekompetenz und Mathematikkompetenz zeigen sich auch empirisch in vergleichsweise hohen (messfehlerbereinigten) Korrelationen zwischen beiden Kompetenzbereichen in PISA 2003 (je nach zugrunde gelegter Stichprobe:  $r = .89$ , OECD, 2005, Kap. 13; bzw.  $r = .90$ , Leutner et al., 2004).

Der in PISA 2003 eingesetzte Test zur Erfassung der naturwissenschaftlichen Kompetenz basiert auf dem Konzept der *Scientific Literacy* (z. B. Bybee, 1997) und erfasst, die Fähigkeit „naturwissenschaftliches Wissen anzuwenden, naturwissenschaftliche Fragen zu erkennen und aus Belegen Schlussfolgerungen zu ziehen, um Entscheidungen zu verstehen und zu treffen [...]“ (OECD, 2003, S. 133; vgl. Rost, Walter, Carstensen, Senkbeil & Prenzel, 2004). Auch für diesen Test zeigen sich theoretisch-konzeptuelle Ähnlichkeiten mit der analytischen Problemlösekompetenz, wie Rumann et al. (2010) anhand einer Analyse der entsprechenden PISA-Testaufgaben zeigen konnten, die sich ebenfalls empirisch in vergleichsweise hohen (messfehlerbereinigten) Korrelationen dieser Kompetenzskalen niederschlagen (je nach zugrunde gelegter Stichprobe:  $r = .78$ , OECD, 2005, Kap. 13; bzw.  $r = .85$ , Leutner et al., 2004).

Für die analytische Problemlösekompetenz zeigte sich in Deutschland ein deutlicher Niveauunterschied: Die (im internationalen Vergleich hohe) analytische Problemlösekompetenz der Schülerinnen und Schüler übersteigt die (im internationalen Vergleich durchschnittliche) mathematische Kompetenz um zehn Punkte auf der internationalen PISA-Metrik. Dieses Ergebnis wird im internationalen und im nationalen Berichtsband zu PISA 2003 (Leutner et al., 2004; OECD, 2004, Kap. 4) als mangelnde Ausschöpfung des beim Problemlösen offensichtlich werdenden kognitiven Potenzials zum Aufbau fachbezogener Kompetenzen interpretiert (Potenzialausschöpfungshypothese). Ähnliches lässt sich für den Vergleich der analytischen Problemlösekompetenz mit der (im internationalen Vergleich durchschnittlichen) naturwissenschaftlichen Kompetenz zeigen. Hier beträgt die Differenz zugunsten der Problemlösekompetenz elf Punkte (Rost et al., 2004).

### 1.3.4 Validität und Kompetenzmessung

Fächerübergreifendes analytisches Problemlösen wurde häufig im Rahmen der Wissenspsychologie und der Expertiseforschung untersucht, sodass die Messung von Unterschieden zwischen Personen und Fragen der Validität nicht im Fokus standen (z. B. Chase & Simon, 1973; Holyoak, 1991; Larkin, McDermott, Simon & Simon, 1980; Sweller, Mawer & Mard, 1983). Interindividuelle Unterschiede im fächerübergreifenden analytischen Problemlösen von Schülerinnen und Schülern im Sinne von Kompetenzunterschieden wurden

im Vergleich zu Unterschieden in fachbezogenen Kompetenzen in der Forschung bisher seltener thematisiert. Bei der Forschung zur Modellierung von Kompetenzen stehen allgemein Fragen nach geeigneten theoretischen und psychometrischen Modellen von Kompetenzen und Fragen der Operationalisierung im Zentrum des Interesses (Fleischer et al., 2013; Hartig, Klieme & Leutner, 2008). Damit rückt auch die Untersuchung von Aspekten der Validität von Kompetenzmodellierungen in den Fokus (z. B. Leuders, 2014; Leutner, Klieme, Fleischer & Kuper, 2013; Tiffin-Richards & Pant, 2017).

Unter Validität, dem zentralen Testgütekriterium der klassischen Testtheorie, wird häufig, etwas vereinfacht, eine Antwort auf die Frage verstanden, ob ein Test das misst, was er messen soll, beziehungsweise was er zu messen vorgibt (z. B. Döring & Bortz, 2016). Umfassender definiert Messick (1989) Validität als „an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment“ (Messick, 1989 S. 13). Während bis zur Mitte des 20. Jahrhunderts Validität vornehmlich als Eigenschaft eines Testinstruments angesehen wurde (APA, 1954), wird diese mittlerweile vor allem auf die Verwendung und Interpretation der Testergebnisse bezogen (AERA, APA & NCME, 1999). Borsboom, Mellenbergh und van Heerden (2004) propagieren jedoch eine engere Fassung des Validitätsbegriffs, bei der die Validierung des Testinstruments im Vordergrund steht. Ein Test ist demnach valide, wenn (a) das zu messende Merkmal tatsächlich existiert und (b) Variation des Merkmals Variation der Test-scores produziert, diese also kausal bedingt. Auch wenn diese Konzeption, ebenso wie wesentlich umfassendere Konzeptionen von Validität (z. B. Kane, 2013), nicht unumstritten ist, kann sie als Leitfaden für den Forschungsprozess dienen, da sie hilft, den Prozess der Konstrukt- und Testvalidierung auf die wesentlichen Aspekte zu fokussieren.

Neben der *Inhaltsvalidität*, also der Frage, inwiefern die Items eines Tests das zu messende Merkmal tatsächlich adäquat repräsentieren (Hartig, Frey & Jude, 2012; siehe Baumert, Stanat & Demmrich, 2001, zur Frage der curricularen Validität der bei PISA 2000 eingesetzten Testverfahren) stand in der ersten Hälfte des 20. Jahrhunderts die *Kriteriumsvalidität* in der psychologischen Leistungsdiagnostik im Vordergrund. Kriteriumsvalidität bezieht sich auf die Frage, inwiefern von einem Testergebnis auf ein für diagnostische Entscheidungen relevantes Außenkriterium geschlossen werden kann (z. B. Cronbach &

Gleser, 1965). In der Folge rückten Fragen der *Konstruktvalidität* stärker in den Fokus, die inzwischen als zentrales Validitätskonzept verstanden wird, unter das alle anderen Arten der Validität subsumiert werden können (Anastasi, 1986; Embretson, 1983; Messick, 1995). Kern der Konstruktvalidität ist die Frage, wie Testergebnisse vor dem Hintergrund eines theoretischen Konstrukts zu interpretieren sind (Messick, 1995). Ursprünglich stand im Rahmen der Untersuchung der Konstruktvalidität der Bezug auf eine formalisierte Theorie und daraus abgeleiteter Hypothesen im Zentrum (z. B. Cronbach & Meehl, 1955). Demgegenüber wird heutzutage bei der Validierung der Interpretation eines Testergebnisses, häufig in Ermangelung entsprechend formalisierter Theorien, sehr viel breiter nach theoretisch fundierten und empirisch belegten Argumenten für und wider eine spezifische Testwertinterpretation gesucht (AERA, APA & NCME, 1999, Kane, 1992, 2001; vgl. Borsboom et al., 2004). Messick (1995) unterscheidet in seinem Modell sechs miteinander verbundene und sich teilweise überlappende Aspekte beziehungsweise Standards zur Prüfung der Konstruktvalidität: (1) *inhaltlicher Aspekt*, (2) *substanzieller oder kognitiver Aspekt*, (3) *struktureller Aspekt*, (4) *Generalisierbarkeit*, (5) *externer Aspekt* und (6) *konsequenzieller Aspekt*. Von diesen sind zwei Aspekte für die empirischen Studien der vorliegenden Arbeit zentral und werden im Folgenden dargestellt.

*Struktureller Aspekt:* Dieser Aspekt bezieht sich auf die Frage der Binnenstruktur und der Dimensionalität (faktorielle Validität) des zu untersuchenden Konstrukts. Im Rahmen der Modellierung von Kompetenzen im Bildungsbereich wurde die Entwicklung von Kompetenzmodellen und geeigneten Messinstrumenten als Forschungsdesiderat erkannt (z. B. Hartig et al., 2008; Koeppen et al., 2008). In der Folge wurde verstärkt Augenmerk auf die empirische Überprüfung von Kompetenzmodellen, vor allem Kompetenzstrukturmodellen und teilweise auch Kompetenzentwicklungsmodellen, gelegt (z. B. Artelt & Schneider, 2011; Fleischer, Leutner & Klieme, 2012; Leutner, Fleischer, Grünkorn & Klieme, 2017; Leutner et al., 2013). Kompetenzstrukturmodelle befassen sich mit der Frage, in welche und wie viele verschiedene Dimensionen eine bestimmte Kompetenz differenziert werden kann. Diese Dimensionen repräsentieren inhaltlich unterscheidbare Aspekte, anhand derer sich Kompetenzunterschiede zwischen Personen beschreiben lassen. Die Binnenstruktur von Kompetenzen ist dabei vornehmlich durch die in einem bestimmten Bereich zu bewäl-

tigenden Anforderungen und die hierfür als notwendig erachteten kognitiven Prozesse bedingt (Hartig & Klieme 2006). Die Untersuchung von Aspekten der faktoriellen Validität der Modellierung analytischer Problemlösekompetenz ist Gegenstand von Studie 1 (siehe Kapitel 2) der vorliegenden Arbeit.

*Externer Aspekt:* Hierunter fallen zum einen Fragen nach Zusammenhängen mit sowohl theoretisch verwandten als auch mit theoretisch unterschiedlichen Konstrukten im Sinne der *konvergenten* und *diskriminanten* Validität (Campbell & Fiske, 1959). Zum anderen lassen sich hierunter auch Fragen nach Zusammenhängen mit relevanten zeitgleich erhobenen Außenkriterien (*Übereinstimmungsvalidität*) sowie relevanten zeitlich später erfassten Außenkriterien (*prognostische Validität*) im Sinne der *Kriteriumsvalidität* fassen (Cronbach & Gleser, 1965; siehe auch Hartig, Frey & Jude, 2012). Darüber hinaus lassen sich unter diesem Aspekt auch Fragen der Nützlichkeit einer Testanwendung fassen, beispielsweise wenn es um die Frage geht, ob ein Testinstrument zusätzliche diagnostisch relevante Informationen über ein anderes Testinstrument hinaus liefert. Diese Aspekte der Validität der Modellierung analytischer Problemlösekompetenz sind Gegenstand der Studien 2 und 3 (siehe Kapitel 3 und 4) der vorliegenden Arbeit.

Eine weitere eher angewandte Form der Validität und gleichzeitig eine sehr direkte und stringente Form der Validitätsprüfung stellt die *inkrementelle Validität* dar, bei der geprüft wird, ob die Vorhersage eines relevanten Kriteriums durch die Hinzunahme eines weiteren Testinstruments gegenüber einem bereits existierenden Instrument verbessert werden kann (Haynes & Lench, 2003; Sechrest, 1963). Die inkrementelle Validität kann als eigenständige Validitätsart betrachtet werden (z. B. Haynes & Lench, 2003) oder aber auch als Aspekt der Nützlichkeit eines Testinstruments beziehungsweise Konstrukts und damit unter den externen Aspekt der Konstruktvalidität (sensu Messick, 1995) subsumiert werden (z. B. Hunsley, 1987; Hunsley & Meyer, 2003). Dieser Aspekten der Validität der Modellierung analytischer Problemlösekompetenz wird in Studie 2 und insbesondere in Studie 3 der vorliegenden Arbeit untersucht.

## 1.4 Struktur und Forschungsfragen

Während die Untersuchung analytischen Problemlösens in der Wissenspsychologie und der Expertiseforschung eine lange Tradition hat, bei der Fragen der Validität jedoch nicht im Vordergrund stehen, wurde analytisches Problemlösen im Rahmen der Forschung zur Modellierung von Kompetenzen, bei der Fragen der Validität besonders im Fokus stehen, bisher seltener betrachtet. Die umfassende Untersuchung von Aspekten der Validität der Modellierung analytischer Problemlösekompetenz stellt somit ein Forschungsdesiderat dar, welches vor dem Hintergrund der zuvor dargestellten Ergebnisse aus PISA 2003, wonach sich beim analytischen Problemlösen kognitives Potenzial manifestiert, welches in der Schule nicht hinreichend zum Aufbau fachbezogener Kompetenzen genutzt wird (Potenzialausschöpfungshypothese), eine besondere Relevanz bekommt. Um dieses Potenzial für didaktische Maßnahmen zum Aufbau fachbezogener Kompetenzen nutzbar zu machen, sind unter anderem Fragen der Binnenstruktur, der Abgrenzbarkeit von anderen schulleistungsrelevanten Konstrukten sowie Fragen nach der prädiktiven Kraft der analytischen Problemlösekompetenz für zukünftige fachbezogene Kompetenzen zu klären. Damit werden zentrale Aspekte der Validität angesprochen, die für die Modellierung fächerübergreifender analytischer Problemlösekompetenz eingehender untersucht werden müssen, als dies in der bisherigen Forschung geschehen ist. Darüber hinaus sollten ergänzende Erklärungen für die Potenzialausschöpfungshypothese geprüft, sowie zusätzlich, die Bedeutung motivationaler und emotionaler Faktoren für erfolgreiches Problemlösen untersucht werden, die – wie zuvor dargestellt – nicht Teil der Kompetenzdefinition sind.

In Studie 1 (Kapitel 2) werden zunächst Aspekte der *faktoriellen Validität* der Modellierung analytischer Problemlösekompetenz untersucht, indem zum einen geprüft wird, inwiefern die drei unterschiedlichen Arten von Problemstellungen – nämlich (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* sowie (3) *Fehler suchen* – empirisch voneinander trennbare Dimensionen abbilden. Zum anderen wird geprüft, inwiefern eine derartige dreidimensionale Modellierung differenziertere diagnostische Informationen liefert, als dies bei der Verwendung aggregierter Informationen aus einer gemeinsamen eindimensionalen Modellierung der Fall ist, was die praktische Bedeutsamkeit dieser Modellierung unterstreichen würde.

In Studie 2 (Kapitel 3) werden Aspekte der *diskriminanten Validität* der Modellierung analytischer Problemlösekompetenz gegenüber fluiden Fähigkeiten als einem zentralen Konzept der Intelligenz untersucht. Hierzu wird geprüft, ob ein Nested-Faktor-Modell, welches von der multiplen Bedingtheit von analytischen Problemlöseleistungen durch analytische Problemlösekompetenz und fluide Fähigkeiten ausgeht, eine bessere Modellpassung aufweist als das Standardmodell, welches analytische Problemlöseleistungen und Leistungen in Items zur Erfassung fluider Fähigkeiten auf jeweils unterschiedliche Faktoren zurückführt. Zusätzlich wird geprüft, ob ein um fluide Fähigkeiten bereinigter Faktor der analytischen Problemlösekompetenz substantiell und inhaltlich interpretierbar ist und inwiefern dieser Faktor praktische Bedeutsamkeit besitzt. Dazu wird geprüft, inwiefern dieser Faktor differenziertere diagnostische Informationen liefert als die Standardmodellierung der analytischen Problemlösekompetenz.

In Studie 3 (Kapitel 4) werden Aspekte der *prognostischen und inkrementellen Validität* der Modellierung analytischer Problemlösekompetenz untersucht. Hierzu wird geprüft, ob analytische Problemlösekompetenz zukünftige fachbezogene Kompetenzen in der Mathematik und den Naturwissenschaften vorhersagen kann, und ob sie außerdem geeignet ist, die Kompetenzentwicklung in diesen Domänen vorherzusagen, bei gleichzeitiger Kontrolle der fluiden Fähigkeiten. Hinsichtlich der inkrementellen Validität wird geprüft, ob analytische Problemlösekompetenz, über den Effekt fluider Fähigkeiten hinaus, inkrementelle Varianz zukünftiger fachbezogener Kompetenzen aufklärt. Zudem wird geprüft, ob analytische Problemlösekompetenz, über die Effekte der jeweiligen Ausgangskompetenzen hinaus, inkrementelle Varianz zukünftiger fachbezogener Kompetenzen aufklärt. Ferner wird geprüft, ob sie auch bei der Vorhersage der Kompetenzentwicklung fachbezogener Kompetenzen einen inkrementellen Beitrag gegenüber fluiden Fähigkeiten leistet.

Studie 4 (Kapitel 5) erweitert die Perspektive der Studien 1 bis 3, indem zusätzlich zu kognitiven Aspekten, auch die Bedeutung motivationaler und emotionaler Faktoren für die analytische Problemlösekompetenz untersucht wird. Es wird untersucht, ob in Ergänzung zur Potenzialausschöpfungshypothese auch von einer mangelnden Nutzung tatsächlich erworbener und vorhandener fachbezogener Kompetenzen bei der Bearbeitung von PISA-Testaufgaben ausgegangen werden kann. Hierfür wird geprüft, inwiefern unterschiedliche



kontextuelle Einbettungen von Problemlöseaufgaben einen Einfluss auf die Schülerleistungen in diesen Aufgaben ausüben können und inwiefern diese Effekte durch das mathematische Selbstkonzept und die Mathematikangst (als leistungsrelevante motivationale, beziehungsweise emotionale, Faktoren) moderiert werden.

## 1.5 Literatur

- AAAS. (1993). Benchmarks for science literacy. New York, NY: Oxford University Press.
- AERA., APA. & NCME. (1999). Standards for educational and psychological testing. Washington, DC: APA.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369–406. doi: 10.1037/0033-295X.89.4.369
- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Reviews of Psychology*, 37, 1–15. doi: 10.1146/annurev.ps.37.020186.000245
- APA. (1954). Technical recommendations for psychological tests and diagnostic techniques. Washington, DC: Author.
- Arbinger, R. (1997). *Psychologie des Problemlösens. Eine anwendungsorientierte Einführung*. Darmstadt: Primus.
- Arlin, P. K. (1989). The problem of the problem. In J. D. Sinnott (Ed.), *Everyday problem solving: Theory and applications* (pp. 229–237). New York, NY: Praeger.
- Artelt, C. & Schneider, W. (2011). Herausforderungen und Möglichkeiten der Diagnose und Modellierung von Kompetenzen und ihrer Entwicklung. Editorial. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 167–172. doi: 10.1026/0049-8637/a000050
- Artzt, A. F. & Armour-Thomas, E. (1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving. *Cognition and Instruction*, 9, 137–175. doi: 10.1207/s1532690xci0902\_3
- Atwood, M. E. & Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology*, 8, 191-216. doi: 10.1016/0010-0285(76)90023-2
- Autor, D., H., Levy, F. & Murnane, R.J. (2003). The skill content of recent technological change: an empirical investigation. *The Quarterly Journal of Economics*, 118, 1279–1333. doi: 10.1162/003355303322552801
- Bandura, A. (1997). *Self-Efficacy – The exercise of control*. New York, NY: Freeman.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium

- (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.  
doi: 10.1007/978 3-322-83412-6\_3
- Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L. & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 205–223.  
doi: 10.1037/0278-7393.21.1.205.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M. & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGraw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Dordrecht: Springer.  
doi: 10.1007/978-94-007-2324-5\_2
- Blomhøj, M. & Hojgaard Jensen, T. (2003). Developing mathematical modelling competence: Conceptual clarification and educational planning. Teaching. *Mathematics and its applications*, 22(3), 123–139. doi: 10.1093/teamat/22.3.123.
- Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (Hrsg.). (2006). *Bildungsstandards Mathematik: konkret*. Berlin: Cornelsen.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, C. H. (2004). Mathematische Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47–92). Münster: Waxmann.
- Blum, W. & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects - state, trend und issues in mathematics instruction. *Educational Studies in Mathematics*, 22, 37–68. doi: 10.1007/BF00302716
- Bodner, G. M. & Herron, J. D. (2002). Problem-solving in chemistry. In J. K. Gilbert, O. de Jong, R. Justi, D. F. Treagust & J. H. van Driel (Eds.), *Chemical education: Towards research-based practise* (pp. 235–266). Dordrecht: Kluwer.
- Bouffard-Bouchard, T., Parent, S. & Larivée, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. *International Journal of Behavioral Development*, 14, 153–164.  
doi: 10.1177/016502549101400203

- Borsboom, D., Mellenbergh, G. J. & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071. doi: 10.1037/0033-295X.111.4.1061
- Bransford, J. D. & Stein, B. S. (1984). *The IDEAL problem solver: A guide for improving thinking, learning and creativity*. New York: Freeman.
- Bunce, D. M. & Heikkinen, H. (1986). The effects of an explicit problem-solving approach on mathematical chemistry achievement. *Journal of Research in Science Teaching*, 23, 11–20. doi: 10.1002/tea.3660230102
- Bybee, R.W. (1997). Toward an understanding of scientific literacy. In: W. Gräber & C. Bolte (Eds.), *Scientific literacy* (pp. 37–68). Kiel: IPN.
- Bybee, R. W. (2000). Teaching science as inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiry into inquiry learning and teaching in science* (pp. 20–46). Washington, DC: AAAS.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. doi: 10.1037/h0046016
- Carlson, M. P. & Bloom, I. (2005). The cyclic nature of problem solving: An emergent multi-dimensional problem-solving framework. *Educational Studies in Mathematics*, 58, 45–75. doi: 10.1007/s10649-005-0808-x
- Chase, W. G. & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81. doi: 10.1016/0010-0285(73)90004-2
- Chi, M. T. H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in psychology of human intelligence* (Vol. 1, pp. 7–75). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press. doi: 10.1037/h0040957
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. doi: 10.1037/h0040957
- Csapó, B. (2004). Knowledge and competencies. In J. Letschert (Ed.), *The integrated person. How curriculum development relates to new competencies* (pp. 35–49). Enschede: CIDREE/SLO.

- Davidson, J. E., Deuser, R. & Sternberg, R. J. (1994). The role of metacognition in problem solving. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 207–226). Cambridge, MA: MIT Press.
- Davis, S. P. (2000). Move evaluation as a predictor and moderator of success in solutions to well-structured problems. *The Quarterly Journal of Experimental Psychology*, 53, 1186–1201. doi: 10.1080/713755943
- Dewey, J. (1910). *How we think*. Boston: Heath.
- Dochy, F. J. R. C. & Alexander, P. A. (1995). Mapping prior knowledge: A framework for discussion among researchers. *European Journal of Psychology of Education*, 10, 225–242. doi: 10.1007/BF03172918
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. überarb. und erw. Aufl.). Berlin: Springer. doi: 10.1007/978-3-642-41089-5
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.
- Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *The nature of insight* (pp. 365–395). Cambridge, MA: MIT Press.
- Dunbar, K. & Fugelsang, J. (2005). Scientific thinking and reasoning. In K. J. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking & reasoning* (pp. 705–725). Cambridge, MA: Cambridge University Press.
- Duncker, K. (1935). *The psychology of productive thinking*. Berlin: Springer.
- Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179–197. doi: 10.1037/0033-2909.93.1.179
- Fleischer, J., Koeppen, K., Kenk, M., Klieme, E. & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Zeitschrift für Erziehungswissenschaft*, 16(Suppl. 1), 5–22. doi:10.1007/s11618-013-0379-z

- Fleischer, J., Leutner, D. & Klieme, E. (Hrsg.). (2012). Modellierung von Kompetenzen im Bereich der Bildung: Eine psychologische Perspektive [Themenheft]. *Psychologische Rundschau*, 63(1). doi: 10.1026/0033-3042/a000111.
- Fleischer, J., Wirth, J., Rumann, S. & Leutner, D. (2010). Strukturen fächerübergreifender und fachlicher Problemlösekompetenz – Analyse von Aufgabenprofilen. *Zeitschrift für Pädagogik*, 56. Beiheft, 239–248.
- Frensch, P. A. & Funke, J. (1995). Definitions, traditions and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving. The european perspective* (pp. 3–25). Hillsdale, NJ: Erlbaum.
- Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J. & Frensch, P. A. (2007). Complex problem solving: The European perspective – 10 years after. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 25–47). New York, NY: Erlbaum.
- Gabel, D. L. & Bunce, D. M. (1994). Research on problem solving: Chemistry. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 301–326). New York, NY: Macmillan.
- Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist*, 21, 99–120. doi: 10.1080/00461520.1986.9653026.
- Greiff, S. & Funke, J. (2017). Interactive problem solving: Exploring the potential of minimal complex systems. In B. Csapó, & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 93–105). Paris: OECD Publishing. doi: 10.1787/20769679
- Greiff, S., Wüstenberg, S. & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, 36, 189–213. doi: 10.1177/0146621612439620
- Harms, U., Mayer, J., Hammann, M., Bayrhuber H. & Kattmann, U. (2004). Kerncurriculum und Standards für den Biologieunterricht in der gymnasialen Oberstufe. In H.-E. Tenorth (Ed.), *Kerncurriculum Oberstufe II* (pp. 22–84). Weinheim: Beltz.

- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Berlin: Springer. doi: 10.1007/978-3-642-20072-4\_7
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer. doi:10.1007/3-540-33020-8\_9
- Hartig, E. Klieme & D. Leutner (Eds.). (2008). *Assessment of competencies in educational contexts*. Göttingen: Hogrefe.
- Hayes, J. R. (1981). *The complete problem solver*. Philadelphia, PA: The Franklin Institute Press.
- Haynes, S. N. & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, 15, 456–466. doi: 10.1037/1040-3590.15.4.456
- Holyoak, K. J. (1991). Symbolic connectionism: Toward third-generation theories of expertise: prospects and limits. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 301–336). Cambridge: Cambridge University Press.
- Hunsley, J. (1987). Cognitive processes in mathematics anxiety and test anxiety: The role of appraisals, internal dialogue, and attributions. *Journal of Educational Psychology*, 79, 388–392.
- Hunsley, J. & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological and statistical issues. *Psychological Assessment*, 15(4), 446–455. doi: 10.1037/1040-3590.15.4.446
- Hussy, W. (1998). *Denken und Problemlösen* (2. überarb. und erw. Aufl.). Stuttgart: Kohlhammer.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology: Research & Development*, 48, 63–85. doi: 10.1007/BF02300500
- Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–535. doi: 10.1037/0033-2909.112.3.527
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38, 319–342. doi: 10.1111/j.1745-3984.2001.tb01130.x

- Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. doi: 10.1111/jedm.12000
- Kintsch, W. & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92, 109–129. doi: 10.1037/0033-295X.92.1.109
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1–48. doi: 10.1207/s15516709cog1201\_1
- Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, 56, 10–13.
- Klieme, E., Funke, J., Leutner, D., Reimann, P. & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47, 179–200. doi: 10.11588/heidok.00008232.
- Klieme, E., Hartig, J. & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Göttingen: Hogrefe.
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52, 876–903.
- Klieme, E., Leutner, D. & Wirth, J. (Hrsg.). (2005). *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie*. Wiesbaden: VS. doi: 10.1007/978-3-322-85144-4
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 139–190). Opladen: Leske + Buderich. doi: 10.1007/978-3-322-83412-6\_5
- KMK. (2005a). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den mittleren Schulabschluss (Beschluss von 16. Dezember 2004)*. München: Wolters Kluwer.



- KMK. (2005b). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Chemie für den mittleren Schulabschluss (Beschluss von 16. Dezember 2004)*. München: Wolters Kluwer.
- KMK. (2005c). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Physik für den mittleren Schulabschluss (Beschluss von 16. Dezember 2004)*. München: Wolters Kluwer.
- Knoblich, G. & Öllinger, M. (2006). Einsicht und Umstrukturierung beim Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen* (S. 1–85). Göttingen: Hogrefe.
- Köller, O. (2009). Bildungsstandards. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 529–548). Wiesbaden: VS.
- Koeppen, K., Hartig, J., Klieme, E. & Leutner, D. (2008). Current issues in competence modelling and assessment. *Zeitschrift für Psychologie / Journal of Psychology*, 216, 61–73. doi: 10.1027/0044-3409.216.2.61
- Larkin, J. H., McDermott, J., Simon, D. P. & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335–1342. doi: 10.1126/science.208.4450.1335
- Leiß, D. & Blum, W. (2006). Beschreibung mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik: konkret* (S. 33–50). Berlin: Cornelsen.
- Lester, F. K. & Kehle, P. E. (2003). From problem solving to modelling: the evolution of thinking about research on complex mathematical activity. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism. Models and modelling perspectives on mathematics problem solving, learning and teaching* (pp. 501–517). Mahwah, NJ: Erlbaum.
- Leuders, T. (2014). Modellierungen mathematischer Kompetenzen – Kriterien für eine Validitätsprüfung aus fachdidaktischer Sicht. *Journal für Mathematik-Didaktik*, 35, 7–48. doi: 10.1007/s13138-013-0060-3
- Leutner, D., Fleischer, J., Grünkorn, J. & Klieme, E. (Eds.). (2017). *Competence assessment in education: Research, models and instruments*. Berlin: Springer. doi: 10.1007/978-3-319-50030-0

- Leutner, D., Funke, J., Klieme, E. & Wirth, J. (2005). Problemlösefähigkeit als fächerübergreifende Kompetenz. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 11–19). Wiesbaden: VS. doi: 10.1007/978-3-322-85144-4\_2
- Leutner, D., Klieme, E., Fleischer, J & Kuper, H. (Hrsg.). (2013) Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Erziehungswissenschaft*, 16(Suppl. 1). doi:10.1007/s11618-013-0379-z
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147–175). Münster: Waxmann.
- Lorenzo, M. (2005). The development, implementation, and evaluation of a problem solving heuristic. *International Journal of Science and Mathematics Education*, 3, 33–58. doi: 10.1007/s10763-004-8359-7
- Maloney, D. P. (1994). Research on problem solving: Physics. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 327–354). New York, NY: Macmillan.
- Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology*, 84, 35–42. doi: 10.1037/0022-0663.84.1.35
- Masters, G. N. (1982). A Rasch Model for partial credit scoring. *Psychometrika*, 47, 149–174. doi: 10.1007/BF02296272
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Handbuch der Theorien in der biologiedidaktischen Forschung* (S. 177–186). Berlin: Springer. doi: 10.1007/978-3-540-68166-3\_16
- Mayer, R. E. (1990). Problem solving. In M. W. Eysenck (Ed.), *The Blackwell dictionary of cognitive psychology* (pp. 284–288). Oxford: Basil.
- Mayer, R. E. (1992). *Thinking, problem solving, cognition*. New York, NY: Freeman.
- Mayer, R. E. (2010). Problem solving and reasoning. In V. G. Aukrust (Eds.), *Learning and cognition in education* (pp. 112–117). Amsterdam: Elsevier.

- Mayer, R. E. & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 411–452). Mahwah, NJ: Erlbaum.
- McCarthy, J. (1956). The inversion of functions defined by turing machines. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies* (pp. 177–181). Princeton, NJ: Princeton University Press.
- Melle, I., Parchmann, I. & Sumfleth, E (2004). Kerncurriculum Chemie – Ziele, Rahmenbedingungen und Ansatzpunkte. In H.-E. Tenorth (Hrsg.), *Kerncurriculum Oberstufe II* (S. 85–147). Weinheim: Beltz.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13–103). New York, NY: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. doi: 10.1037/0003-066X.50.9.741
- Mettes, C. T. C. W., Pilot, A., Roossink, H. J. & Kramers-Pals, H. (1980). Teaching and learning problem solving in science. Part I: A general strategy. *Journal of Chemical Education*, 57, 882–885. doi: 10.1021/ed057p882
- MSJK. (Hrsg.). (2004). *Kernlehrplan für die Gesamtschule– Sekundarstufe I in Nordrhein-Westfalen – Mathematik*. Frechen: Ritterbach (jeweils auch für das Gymnasium, die Hauptschule und die Realschule).
- NCTM. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N. et al. (2001). The German addition to the OECD-PISA mathematics assessment: Framework for the supplementary test and its connection to the international framework. *International Reviews on Mathematical Education*, 33, 45–59. doi: 10.1007/BF02652739
- Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

- OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing.
- OECD. (2004). *Problem solving for tomorrow's world: First measurements of cross-curricular competencies from PISA 2003*. Paris: OECD Publishing.
- OECD. (2005). *PISA 2003: Technical report*. Paris: OECD Publishing.
- OECD (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing.  
doi: 10.1787/9789264190511-en
- OECD. (2014). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems (Volume V)*. Paris: OECD Publishing.
- O'Neil, H. F. Jr. (1999). Perspectives on computer-based performance assessment of problem solving. *Computers in Human Behavior*, 15, 255–268.  
doi 10.1016/S0747 5632(99)00022-9
- Pant, H. A., Böhme, K. & Köller, O. (2012). Das Kompetenzkonzept der Bildungsstandards und die Entwicklung von Kompetenzstufenmodellen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik* (S. 49–55). Münster: Waxmann.
- Paris, S. G., Lipson, M. Y. & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293–316. doi: 10.1016/0361-476X(83)90018-8.
- Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.
- Ramalingam, D., Philpot, R. & McCrae (2017). The PISA 2012 assessment of problem solving. In B. Csapó, & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 75–91). Paris: OECD Publishing.  
doi: 10.1787/20769679
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: The Danish Institute for Educational Research.

- Reiss, K. & Törner, G. (2007). Problem solving in the mathematics classroom: The German perspective. *ZDM-The International Journal on Mathematics Education*, 39, 431–441. doi: 10.1007/s11858-007-0040-5
- Reitman, W. R (1965). *Cognition and thought: An information processing approach*. New York, NY: Wiley.
- Resnick, L. B. (1983). Toward a cognitive theory of instruction. In S. Paris, G. Olson & H. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 5–38). Hillsdale, NJ: Erlbaum.
- Rittle-Johnson, B., Siegler, R. S. & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93, 346–362. doi: 10.1037/0022-0663.93.2.346
- Robertson, S. I. (2001). *Problem solving*. Philadelphia, PA: Taylor & Francis.
- Rost, J., Walter, O., Carstensen, C. H., Senkbeil, M. & Prenzel, M. (2004). Naturwissenschaftliche Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 111–146). Münster: Waxmann.
- Rumann, S., Fleischer, J., Stawitz, H., Wirth, J. & Leutner, D. (2010). Vergleich von Profilen der Naturwissenschafts- und Problemlöse-Aufgaben der PISA 2003-Studie. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 315–327.
- Schneider, W., Schlagmüller, M. & Visé, M. (1998). The impact of metamemory and domain-specific knowledge on memory performance. *European Journal of Psychology of Education*, 13, 91–103. doi: 10.1007/BF03172815
- Schoenfeld, A. H. (1983). Episodes and executive decisions in mathematical problem solving. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 345–395). New York, NY: Academic Press.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grows (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). New York, NY: Macmillan.
- Schreiber, B. (1998). *Selbstreguliertes Lernen*. Münster: Waxmann.

- Schupp, H. (1988). Anwendungsorientierter Mathematikunterricht in der Sekundarstufe I zwischen Tradition und neuen Impulsen. *Der Mathematikunterricht*, 34(6), 5–16.
- Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, 23, 153–158. doi: 10.1177/001316446302300113
- Staub, F. C. & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver III, S. Mannes & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 285–305). Hillsdale, NJ: Erlbaum.
- Sternberg, R. J. & Davidson, J. E. (1992). Problem solving. In M. Alkin (Ed.), *Encyclopedia of educational research* (pp. 1037–1045). New York, NY: Macmillan.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement: Issues and Practice*, 14, 29–36. doi: 10.1111/j.1745 3992.1995.tb00865.x
- Sweller, J., Mawer, R. & Mard, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, 112, 639–661. doi 10.1037/0096-3445.112.4.639
- Tiffin-Richards, S. P. & Pant, H. A. (2017). Arguing validity in educational assessment. . In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 315–331). Berlin: Springer. doi: 10.1007/978-3-319-50030-0\_27
- Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios. Taxonomie, Entwicklung, Evaluation*. Lengerich: PSP.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Seattle, WA: Hogrefe.
- Wertheimer, M. (1945). *Productive thinking*. New York, NY: Harper & Row.

- Wilhelm, O. & Nickolaus, R. (2013). Was grenzt das Kompetenzkonzept von etablierten Kategorien wie Fähigkeit, Fertigkeit oder Intelligenz ab? *Zeitschrift für Erziehungswissenschaft*, 16(Suppl. 1), 23–26. doi: 10.1007/s11618-013-0379-z
- Wirth, J. & Klieme, E. (2003). Computer-based assessment of problem solving performances. *Assessment in Education: Principles, Policy & Practice*, 10, 329–345. doi: 10.1080/0969594032000148172
- Zimmerman, B. J., Bandura, A. & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29, 663–676. doi: 10.3102/00028312029003663

## **2 STUDIE 1: FAKTORIELLE VALIDITÄT (DIMENSIONALITÄT) DER MODELLIERUNG ANALYTISCHER PROBLEMLÖSEKOMPETENZ – DIE BEDEUTUNG UNTERSCHIEDLICHER ARTEN VON PROBLEMSTELLUNGEN**

### **2.1 Einleitung**

Die Frage, in wie viele Dimensionen eine Kompetenz differenziert werden kann und wie diese inhaltlich beschrieben werden können, ist ein zentraler Aspekt der Forschung zur Modellierung von Kompetenzen im Bildungsbereich (z. B. Klieme, Leutner & Kenk, 2010; Leutner, Klieme, Fleischer & Kuper, 2013). Unter Dimensionen werden inhaltlich unterscheidbare Aspekte verstanden, anhand derer Kompetenzunterschiede zwischen Personen beschrieben werden können (Hartig & Klieme, 2006). Die Anzahl dieser Dimensionen und deren Zusammenhänge, also die Struktur von Kompetenzen, werden vornehmlich durch die zu bewältigenden Anforderungen und die dafür notwendigen kognitiven Prozesse bedingt (Hartig & Klieme, 2006). So können die Dimensionen einer Kompetenz beispielsweise durch unterschiedliche inhaltliche Einheiten (z. B. Winkelmann, Robitzsch, Stanat & Köller, 2012), unterschiedliche Verarbeitungsprozesse (z. B. Jude et al., 2008), aber auch durch unterschiedliche Repräsentations- und Aufgabenformate (z. B. Bayrhuber, Leuders, Bruder & Wirtz, 2010; Schnotz et al., 2010) definiert sein (vgl. Fleischer, Koeppen, Kenk, Klieme & Leutner, 2013). Das Ziel von Studie 1 ist die Untersuchung dieses mitunter auch als *faktorielle Validität* bezeichneten Aspekts der Konstruktvalidität (Hartig, Frey & Jude, 2012; *struktureller Aspekt* der Konstruktvalidität, sensu Messick, 1995) in Bezug auf unterschiedliche Arten von Problemstellungen zur Erfassung der analytischen Problemlösekompetenz.



## 2.2 Theoretischer Hintergrund – Operationalisierung analytischer Problemlösekompetenz

Zur Operationalisierung der analytischen Problemlösekompetenz im Rahmen standardisierter Testverfahren werden unterschiedliche Arten von Problemstellungen eingesetzt. In der nationalen Ergänzungsstudie zu PISA 2000, in der analytische Problemlösekompetenz erstmalig im Rahmen einer großen international vergleichenden Schulleistungstudie erfasst wurde, kamen sogenannte Projektaufgaben zum Einsatz (für den Einsatz dieses Aufgabenkonzepts im Rahmen der Untersuchung „Aspekte der Lernausgangslage und der Lernentwicklung [LAU]“ in Hamburg, siehe Ebach, Klieme & Hensgen, 1999 sowie Lehmann, Peek, Gänsfuß & Husfeldt, 2002). Bei diesen Aufgaben wird ein Projekt, beispielsweise die Organisation einer Klassenparty, in einzelne Arbeitsschritte gegliedert, und zu diesen Arbeitsschritten werden einzelne Testaufgaben gestellt (Klieme, Hartig & Wirth, 2005). Aufgaben dieses Typs können als *fächerunabhängig* bezeichnet werden, da sie primär auf planendes und kombinatorisches Denken abzielen und hauptsächlich Alltagswissen und kein fachbezogenes Wissen erfordern (Klieme et al., 2005).

Für den internationalen Test zum Problemlösen in PISA 2003 wurden hingegen *fächerverbindende* Aufgaben konstruiert, bei denen Wissensinhalte aus verschiedenen fachlichen Domänen kombiniert werden, beispielsweise Anforderungen aus der Mathematik mit textbezogenem Interpretieren und/oder Prozessen des naturwissenschaftlichen Denkens (Klieme et al., 2005; Leutner, Klieme, Meyer & Wirth, 2004). Ergebnisse von Fleischer, Wirth, Rumann und Leutner (2010) deuten jedoch darauf hin, dass dieser fächerverbindende Charakter der Aufgaben nur bedingt gegeben ist. So konnten die Autoren anhand einer Aufgabenanalyse zeigen, dass die meisten Problemlöseaufgaben keiner fachlichen Domäne und etwa ein Drittel der Aufgaben der Mathematik zuzuordnen sind. Um ein möglichst breites Spektrum der Anforderungen beim Lösen analytischer Probleme abdecken zu können, kamen drei Arten von Problemstellungen zum Einsatz. (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* (Leutner et al., 2004, OECD, 2003, 2004; für weitere Arten von Testaufgaben zur Erfassung analytischer Problemlösekompetenz, die zuvor im Rahmen des PISA-2000-Feldtests erprobt wurden, siehe Leutner, Wirth, Klieme & Funke, 2005). Ein Beispiel für eine Problemstellung der Art

*Entscheidungen treffen* stellt die Testaufgabe „Anschlusszüge“ dar (siehe Abbildung 1.1 in Abschnitt 1.3.3), bei der es darum geht, die schnellste und kostengünstigste Verbindung zwischen zwei U-Bahn-Stationen herauszufinden. Hierzu müssen Informationen aus der Abbildung eines U-Bahn-Liniennetzes sowie Informationen aus dem Text der Aufgabenstellung entnommen und integriert werden. Problemstellungen der Art *Systeme analysieren und entwerfen* lassen sich am Beispiel der Aufgabe „Ferienlager“ verdeutlichen (siehe Abbildung 1.2 in Abschnitt 1.3.3), bei der kombinatorisches Denken im Vordergrund steht. Bei dieser Aufgabe geht es darum, Jungen und Mädchen sowie Erwachsenen verschiedene Schlafsäle zuzuweisen, wobei komplexe, miteinander verbundene einschränkende Randbedingungen hinsichtlich des zahlenmäßigen Verhältnisses Erwachsener zu Kind sowie in Bezug auf die Geschlechterzusammensetzung und die Anzahl der Betten pro Schlafsaal berücksichtigt werden müssen. Ein Beispiel für eine Problemstellung der Art *Fehler suchen* stellt die Aufgabe „Bewässerung“ dar (siehe Abbildung 1.3 in Abschnitt 1.3.3), bei der schlussfolgerndes Denken sowie systematisches Vorgehen zentrale Anforderungen sind. Diese Aufgabe besteht aus mehreren Testitems, bei denen es darum geht, zunächst die Funktionsweise eines Systems von Bewässerungskanälen, welches durch Schleusen gesteuert wird, zu verstehen und im Anschluss verschiedene Fehlfunktionen dieses Systems zu identifizieren. Tabelle 2.1 stellt die Anforderungen der drei Arten von Problemstellungen, wie sie in PISA 2003 zum Einsatz kamen, dar. Wie Tabelle 2.1 zu entnehmen ist, bestehen sowohl Ähnlichkeiten als auch deutliche Unterschiede zwischen den verschiedenen Arten von Problemstellungen hinsichtlich der zu erreichenden Ziele, der beteiligten Prozesse und der potenziell schwierigkeiterzeugenden Merkmale (vgl. Leutner et al., 2004; OECD, 2003, Kap. 4).

**Tabelle 2.1: Ziele, kognitive Prozesse und potenziell schwierigkeiterzeugende Merkmale der drei Arten von Problemstellungen aus PISA 2003**

<b>Art der Problemstellung</b>	<b>Entscheidungen treffen</b>	<b>Systeme analysieren und entwerfen</b>	<b>Fehler suchen</b>
Ziel	Aus einer Liste von Handlungsoptionen unter Berücksichtigung von Bedingungen eine Auswahl treffen	Die Abhängigkeit zwischen Teilen eines Systems identifizieren und ein Umsetzungskonzept zur Bewältigung einer Aufgabenstellung entwickeln	Den Fehler eines nicht angemessen funktionierenden Systems oder Mechanismus diagnostizieren und gegebenenfalls korrigieren
Kognitive Prozesse	Eine Situation verstehen, bei der es zur Erledigung einer spezifischen Aufgabe verschiedene Optionen und Bedingungen zu berücksichtigen gilt	Ein System im Hinblick auf die Anforderungen einer spezifischen Aufgabe verstehen	Die maßgeblichen Eigenschaften eines Systems oder eines Mechanismus und seiner Fehlfunktion im Hinblick auf eine spezifische Aufgabenstellung verstehen
	Relevante Bedingungen erkennen	Relevante Komponenten eines Systems erkennen	Ursächlich aufeinander bezogene Variablen erkennen
	Mögliche Handlungsoptionen zusammenstellen und verstehen	Abhängigkeiten zwischen den Komponenten eines Systems verstehen	Die Funktionsweise des Systems verstehen
	Zwischen Handlungsoptionen entscheiden	Ein System analysieren oder entwerfen, welches die Abhängigkeiten zwischen den Systemkomponenten in angemessener Weise umsetzt	Die Fehlfunktion des Systems diagnostizieren und eine Lösung entwickeln

<b>Art der Problemstellung</b>	<b>Entscheidungen treffen</b>	<b>Systeme analysieren und entwerfen</b>	<b>Fehler suchen</b>
	Die Entscheidung prüfen und bewerten	Die Analyse oder den Entwurf prüfen und bewerten	Die Diagnose und den Lösungsvorschlag prüfen und bewerten
	Die Entscheidung mitteilen und begründen	Die Analyse mitteilen oder den Entwurf begründen	Die Diagnose und den Lösungsvorschlag mitteilen und begründen
Potenziell schwierigkeits-erzeugende Merkmale	Anzahl der Bedingungen	Anzahl der miteinander verknüpften Variablen und die Art der Beziehungen untereinander	Anzahl der voneinander abhängigen Komponenten des Systems oder Mechanismus und die Art der gegenseitigen Abhängigkeit
	Anzahl und Art der verwendeten Darstellungen (verbal, bildlich, numerisch)	Anzahl und Art der verwendeten Darstellungen (verbal, bildlich, numerisch)	Anzahl und Art der verwendeten Darstellungen (verbal, bildlich, numerisch)

*Anmerkungen.* Aus Leutner et al. (2004, S. 150).

Die Entscheidung, fächerverbindende Problemlöseaufgaben in PISA 2003 zu verwenden, mag darin begründet sein, dass diese, wie zuvor dargestellt, einen vermeintlich stärkeren Bezug zu fachlichen Inhalten aufweisen und somit davon auszugehen ist, dass sie sich besser von Maßen der Intelligenz abgrenzen lassen, als dies für fächerunabhängige Aufgaben der Fall ist. Allerdings lässt sich das Konzept der fächerunabhängigen Projektaufgaben aus PISA 2000 auch in den Problemlöseaufgaben der Art *Entscheidungen treffen* und *Systeme analysieren und entwerfen* aus PISA 2003 wiederfinden, die ebenfalls Projektaufgaben darstellen, wobei jedes Projekt hier auf wenige Testitems beschränkt wird (Klieme et al., 2005). Trotz der mitunter deutlichen konzeptuellen Unterschiede der drei verwendeten Arten von Problemstellungen wurde – ebenso wie die internationalen Tests zu den fachbezogenen Kompetenzen Mathematik, Naturwissenschaft und Lesen – auch der Test

zur fächerübergreifenden analytischen Problemlösekompetenz in PISA 2003 als eindimensionales Konstrukt modelliert. Die Leistungen der Schülerinnen und Schüler in den Problemlöseitems wurden somit ohne Differenzierung nach unterschiedlichen Arten von Problemstellungen auf einer gemeinsamen Kompetenzskala erfasst, die dann als Basis für die weiteren Analysen zur analytischen Problemlösekompetenz diente.

## 2.3 Fragestellungen und Hypothesen

Die eindimensionale Skalierung der Problemlösedaten in PISA 2003 erscheint vor dem Hintergrund der Zielsetzung von PISA – der objektiven Messung des Erfolgs von Bildungssystemen im Sinne eines Systemmonitorings (OECD, 2003) – gerechtfertigt, wie auch die im technischen Report zu PISA 2003 berichteten Skalenanalysen zeigen (OECD, 2005). Dennoch bleibt im Sinne des strukturellen Aspekts der Konstruktvalidität (faktorielle Validität) die Frage offen, ob die drei Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* aufgrund unterschiedlicher kognitiver Anforderungen verschiedene empirisch voneinander trennbare Dimensionen abbilden. Die in Tabelle 2.1 dargestellten Unterschiede hinsichtlich der zu erreichenden Ziele, der beteiligten Prozesse und der potenziell schwierigkeiterzeugenden Merkmale lassen dies vermuten. Es wird daher davon ausgegangen, dass eine dreidimensionale Modellierung analytischer Problemlösekompetenz entsprechend den drei Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* die Datenstruktur in PISA 2003 adäquater abbilden kann als eine eindimensionale Modellierung ohne entsprechende Differenzierung (Hypothese 1).

Die Frage, ob ein eindimensionales oder ein mehrdimensionales Messmodell zur Modellierung von Kompetenzdaten verwendet werden sollte, lässt sich jedoch nicht unabhängig von der intendierten Verwendung der Testergebnisse beantworten (Koeppen, Hartig, Klieme & Leutner, 2008; Neumann, 2013). Während, wie bereits angesprochen, im Kontext von PISA aggregierte Informationen einer eindimensionalen Modellierung ihren Zweck durchaus erfüllen können, erfordert die Kompetenzdiagnostik als Spezialfall der pädagogisch-psychologischen Diagnostik in der Regel differenziertere Informationen auf individu-

eller Ebene (Klieme & Leutner, 2006; vgl. Leutner, Fleischer, Spoden & Wirth, 2007). Entscheidender als die Frage nach der optimalen Datenrepräsentation durch ein- oder mehrdimensionale Modelle ist für die diagnostische Praxis somit die Frage, welche Konsequenzen eine Diagnostik auf Basis unterschiedlicher Modellierungen hat. Vor dem Hintergrund der dargestellten Unterschiede hinsichtlich der zu erreichenden Ziele, der beteiligten Prozesse und der potenziell schwierigkeiterzeugenden Merkmale der drei Arten von Problemstellungen wird erwartet, dass sich für die Dimensionen des dreidimensionalen Modells differenzielle Zusammenhänge mit externen Variablen feststellen lassen (Hypothese 2). Das würde bedeuten, dass eine Kompetenzmessung auf Basis von drei getrennten Dimensionen entsprechend den drei Arten von Problemstellungen differenziertere diagnostische Informationen liefern könnte als die Verwendung aggregierter Informationen aus einer gemeinsamen Dimension. Dies würde für die praktische Bedeutsamkeit dieser Modellierung sprechen.

## **2.4 Methode**

### **2.4.1 Stichprobe und Datengrundlage**

Für die Untersuchung wurden die Daten der deutschen PISA-Haupterhebung 2003 zur Problemlösekompetenz ( $N = 2343$ ; Probanden mit Testheften, in denen entsprechende Itemcluster enthalten waren) ausgewählt. Die Stichprobe besteht aus 15-jährigen Schülerinnen und Schülern (51.7 % weiblich), die aus 198 Schulen der Bildungsgänge Hauptschule ( $n = 492$ ), Realschule ( $n = 836$ ), Integrierte Gesamtschule ( $n = 236$ ) und Gymnasium ( $n = 779$ ) rekrutiert wurden. 1960 Schülerinnen und Schüler (83.7 %) gaben als Familiensprache Deutsch an. Für eine detailliertere Beschreibung der in PISA 2003 realisierten Stichprobe und zum Design der Studie sei auf Prenzel, Drechsel, Carstensen und Ramm (2004) sowie OECD (2005) verwiesen.

### **2.4.2 Instrumente**

Zur Erfassung der analytischen Problemlösekompetenz wurden die Rohwerte der 18 Testitems der entsprechenden PISA-Testskala verwendet (PISA-Konsortium Deutschland,

2006, Teil I, Kap. 5). Zur Erfassung der Mathematikkompetenz, der naturwissenschaftlichen Kompetenz und der Lesekompetenz wurden ebenfalls die Daten der entsprechenden PISA-Skalen herangezogen (siehe PISA-Konsortium Deutschland, 2006, Teil I, Kap. 1-3). Als Indikator für Intelligenz wurden darüber hinaus die Daten der Skala *N2 Figurenanalogien* des in PISA eingesetzten *Kognitiven Fähigkeitstests* (KFT 5-12+R; Heller & Perleth, 2000; siehe PISA-Konsortium Deutschland, 2006, Teil I, Kap. 6) verwendet. Diese Subskala erfasst schlussfolgerndes Denken (*reasoning*) beziehungsweise fluide Fähigkeiten, die als Kernkonzepte der Intelligenz betrachtet werden können (z. B. Horn & Noll, 1997; Rost, 2009, Kap. 2; siehe hierzu auch die Ausführungen in Abschnitt 3.2.2). Darüber hinaus wurden die in PISA erhobenen Informationen zu Familiensprache, Geschlecht sowie Bildungsgang in den Analysen verwendet (PISA-Konsortium Deutschland, 2006, Teil II, Kap. 1 & 2).

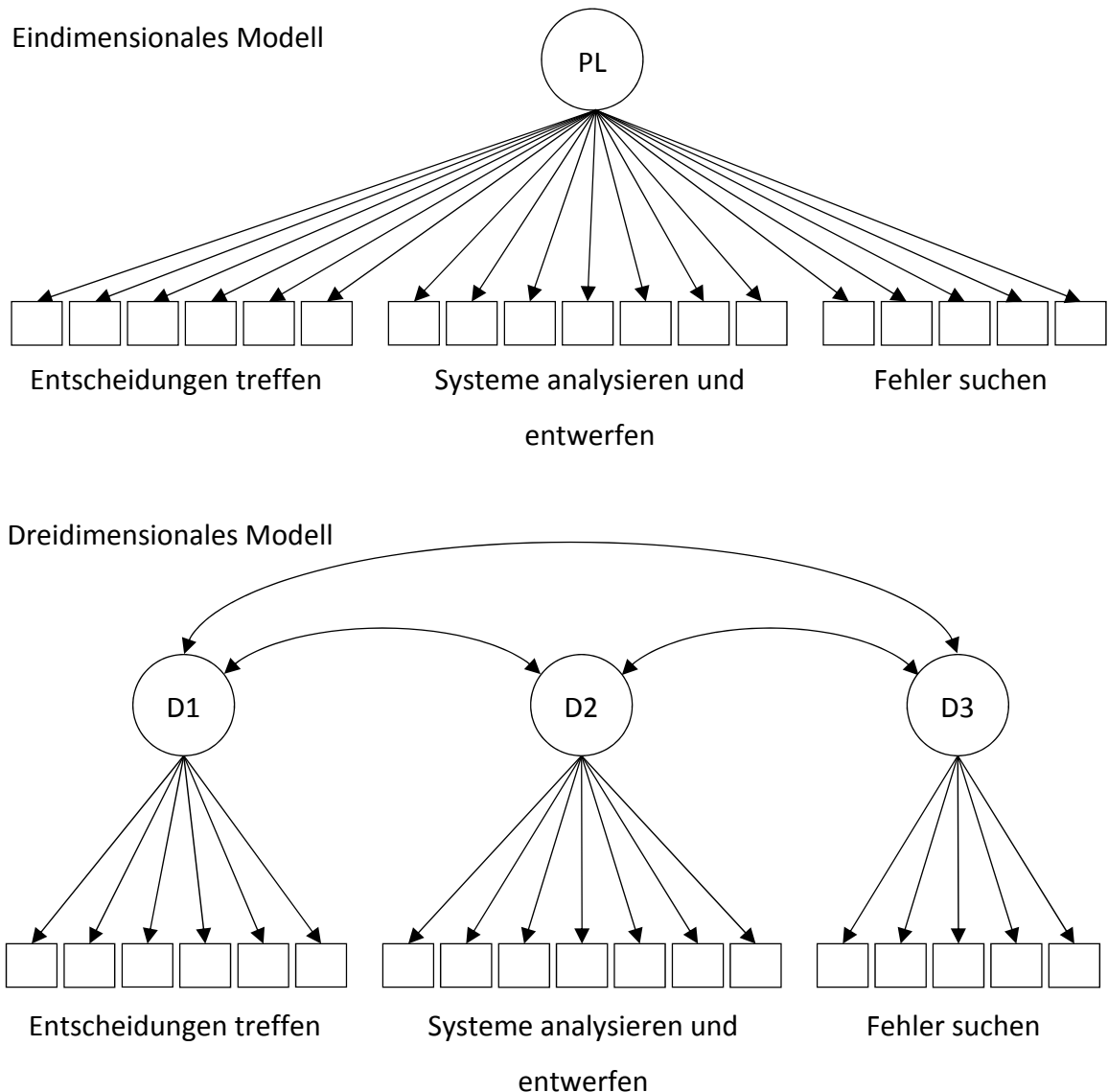
### 2.4.3 Vorgehen und statistische Analysen

Die Daten des Problemlösetests wurden analog zum Vorgehen in PISA 2003 mit dem Mixed-Coefficients-Multinomial-Logit-Modell, wie es im Programm ConQuest 2.0 (Wu, Adams, Wilson & Haldane, 2007) implementiert ist, reskaliert. Die 18 Items des Tests gingen entsprechend der in PISA verwendeten Kodierung entweder als dichotome (0 = *falsch*; 1 = *richtig*) oder polytome Items (0 = *falsch*; 1 = *teilweise richtig*; 2 = *vollständig richtig*) in die Skalierung ein. Das spezifizierte Modell entspricht damit dem Partial-Credit-Modell (Masters, 1982), einer Erweiterung des dichotomen Rasch-Modells (Rasch, 1960) für ordinale Daten. Zunächst wurde ein Modell spezifiziert, welches für das Problemlösen eine Dimension postuliert. Im Anschluss daran wurde ein dreidimensionales Modell spezifiziert, welches für jedes Item in Abhängigkeit von der jeweiligen Art der Problemstellung die Zugehörigkeit zu einer von drei Dimensionen postuliert (*between-item multidimensionality*; Adams, Wilson & Wang, 1997). Die Abbildung 2.1 stellt beide Modelle schematisch dar.

Zur Bewertung beider Modelle wurden verschiedene Kriterien herangezogen, zum einen die Infit-Statistik (weighted MNSQ; siehe Wilson 2005), die über die Abweichungen der beobachteten Itemantworten von den aufgrund des Testmodells erwarteten Itemantworten gebildet wird und die damit eine Information über die Modellpassung jedes Items

liefert (Werte zwischen .75 und 1.33 können als hinreichend guter Itemfit interpretiert werden; Adams & Khoo, 1996; vgl. Wright & Linacre, 1994). Zum anderen wurden die Reliabilitäten und Varianzen der Skalen sowie die Höhe der Korrelationen zwischen den latenten Dimensionen des dreidimensionalen Modells bestimmt. Darüber hinaus wurden das AIC (Akaike Information Criterion) und das BIC (Bayesian Information Criterion) als informationstheoretische Maße der Modellpassung ermittelt (niedrigere Werte stehen jeweils für eine bessere Passung des Modells an die Daten; Kuha, 2004) sowie der LQT (Likelihood-Quotienten-Test) durchgeführt, der einen inferenzstatistischen Vergleich der Güte der Modellpassung verschiedener geschachtelter Modelle erlaubt (Rost, 2004).





**Abbildung 2.1: Eindimensionales und dreidimensionales Modell der analytischen Problemlösekompetenz**

Im Anschluss wurden Zusammenhänge der Personenparameter (WL-Schätzer; Warm, 1989) des dreidimensionalen Modells mit den Variablen Bildungsgang, Familiensprache, Geschlecht, den PISA-Testleistungen in Mathematik, Naturwissenschaften und Lesen sowie fluiden Fähigkeiten untersucht. Die verwendeten Daten weisen eine geclusterte Datenstruktur auf (Schülerinnen und Schüler geclustert in Schulen). Cluster stellen, allgemein gesprochen, Beobachtungseinheiten dar, deren Mitglieder sich aufgrund von Ge-

meinsamkeiten ähneln (z. B. vergleichbarer Lehr- und Lernbedingungen, vergleichbarer soziodemographischer Hintergrundmerkmale etc.; z. B. Hox, 2002). Es ist daher davon auszugehen, dass die Ergebnisse von Schülerinnen und Schülern innerhalb eines Clusters (Schule) ähnlicher sind als die Ergebnisse im Vergleich zwischen den Clustern. Als Maß für die Ähnlichkeit der Beobachtungseinheiten kann die Intraklassenkorrelation (ICC) bestimmt werden, die den Anteil der Gesamtvarianz der Leistungen angibt, der allein auf die Gruppenzugehörigkeit (Schule) der Schülerinnen und Schüler zurückgeführt werden kann (z. B. Goldstein, 2011). Da bereits geringe ICCs zu einer Inflation des alpha-Fehlers bei der Signifikanztestung führen können (z. B. Goldstein, 2011), erfolgten die Analysen mit dem Programm HLM 7 (Raudenbush, Bryk, Cheong, Congdon & du Toit, 2011), wodurch korrekte Schätzungen der Standardfehler der Korrelationskoeffizienten und der Mittelwerte und damit eine korrekte Signifikanzprüfung gewährleistet sind. Die Personenparameter der latenten Dimensionen des dreidimensionalen Modells analytischer Problemlösekompetenz wurden für die Analysen z-standardisiert. In PISA werden als Schätzer der Personenfähigkeiten sogenannte *Plausible Values* verwendet. Hierbei handelt es sich um fünf Zufallsziehungen aus den Wahrscheinlichkeitsverteilungen der Kompetenzen für jede Person, zu deren Schätzung neben den jeweiligen Testdaten auch zusätzliche Hintergrundinformationen über die Schülerinnen und Schüler herangezogen werden und die dadurch unter anderem eine hohe Messgenauigkeit aufweisen (Carstensen, Knoll, Rost & Prenzel, 2004; OECD, 2005, Kap. 9). Für Mathematik, Naturwissenschaften, Lesen und fluide Fähigkeiten wurden diese fünf Plausible Values aus PISA 2003 für jede Skala verwendet. Diese wurden ebenfalls zunächst z-standardisiert. Anschließend wurde für jeden Plausible Value die Korrelation mit den Personenparametern einer Dimension bestimmt, und diese fünf Werte wurden für jede Skala entsprechend gemittelt (zum Hintergrund der Schätzung und Verwendung von Plausible Values siehe z. B. Mislevy, Beaton, Kaplan & Sheehan, 1992; Wu, 2005). Die Variablen Familiensprache und Geschlecht gingen als dummy-kodierte Variablen in die Analysen ein. Die für diese Variablen berichteten Koeffizienten geben daher jeweils die Differenz der geschätzten Mittelwerte zwischen beiden Kategorien an. Die Bildungsgänge wurden ebenfalls dummy-kodiert. Hierbei geben die berichteten Koeffizienten jeweils die Mittelwertdifferenz zur Hauptschule an, die als Referenzkategorie gewählt wurde.

## 2.5 Ergebnisse

Im eindimensionalen Modell zeigten alle Items hinreichend gute Infit-Werte (0.84–1.18). Das Gleiche galt für die Items im dreidimensionalen Modell, wie Tabelle 2.2 zu entnehmen ist. Die Reliabilität des eindimensionalen Modells, erfasst über die EAP-Reliabilität (vgl. Rost, 2004), lag mit  $r = .78$  geringfügig über den Reliabilitäten der Dimensionen des dreidimensionalen Modells (siehe Tabelle 2.2). Die Varianz als Indikator für die Trennschärfe der Items lag für das eindimensionale Modell bei  $\sigma^2 = 1.37$  ( $SE = .064$ ,  $p < .001$ ). Wie Tabelle 2.2 zeigt, bewegte sie sich damit in der gleichen Größenordnung wie die Varianzen der Dimension *Entscheidungen treffen* und *Systeme analysieren und entwerfen* des dreidimensionalen Modells, lag jedoch deutlich unter der Varianz der Dimension *Fehler suchen*. Des Weiteren waren die latenten Dimensionen des dreidimensionalen Modells hoch korreliert. Dies galt insbesondere für die Dimensionen *Entscheidungen treffen* und *Systeme analysieren und entwerfen* (siehe Tabelle 2.2).

**Tabelle 2.2: Reliabilitäten, Interkorrelationen und Varianzen ( $\sigma^2$ ) der latenten Dimensionen des dreidimensionalen Modells der analytischen Problemlösekompetenz sowie Infit-Werte der zugehörigen Testitems**

	D-1	D-2	D-3
Entscheidungen treffen (D-1; 6 Items)	.73		
Systeme analysieren und entwerfen (D-2; 7 Items)	.93	.73	
Fehler suchen (D-3; 5 Items)	.83	.80	.69
$\sigma^2$ (SE)	1.43 (.003)***	1.21 (.004)***	3.60 (.005)***
Infit (weighted MNSQ)	0.98–1.05	0.96–1.12	0.86–1.16

*Anmerkungen.* Reliabilitäten (EAP-Schätzungen) auf der Hauptdiagonalen, Korrelationen der latenten Dimensionen unterhalb der Hauptdiagonalen; \*\*\* $p < .001$ ;  $N = 2343$ .

Tabelle 2.3 stellt die Ergebnisse des LQT sowie die Werte des AIC und BIC für beide spezifizierten Modelle dar. Für das restriktivere eindimensionale Modell zeigte sich eine signifikant schlechtere Passung an die Daten als für das dreidimensionale Modell ( $\Delta\text{Deviance} = 184_{(5)}$ ,  $p < .001$ ). Ebenso sprachen die höheren Werte des AIC und des BIC des eindimensionalen Modells für eine schlechtere Passung dieses Modells im Vergleich zum dreidimensionalen Modell. Hypothese 1 konnte somit bestätigt werden.

**Tabelle 2.3: Modellpassung für das eindimensionale und das dreidimensionale Modell der analytischen Problemlösekompetenz**

	Deviance	Anzahl geschätzter Parameter	$\Delta\text{Deviance}$ ( <i>df</i> )	AIC	BIC
dreidimensionales Modell	31466	32	-	31530	31714
eindimensionales Modell	31650	27	184*** (5)	31704	31859

*Anmerkungen.* Deviance:  $-2\log[\text{LR}]$  der Modellschätzung;  $\Delta\text{Deviance}$ :  $\chi^2$ -verteilte Prüfgröße des Likelihood-Quotienten-Tests mit der Differenz der in beiden Modellen geschätzten Anzahl an Parametern als Freiheitsgrade (*df*); AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; \*\*\* $p < .001$ ;  $N = 2343$ .

Die Ergebnisse der Untersuchung der Zusammenhänge der Personenparameter des dreidimensionalen Modells mit externen Variablen sind in Tabelle 2.4 dargestellt. In allen Dimensionen lagen substantielle Anteile an der jeweiligen Gesamtvarianz der Personenparameter auf Schulebene (erfasst über die ICC), was eine mehrebenenanalytische Auswertung der Daten indiziert. Für die Dimension *Systeme analysieren und entwerfen* lag dieser Anteil mit 32 Prozent jedoch höher als für die anderen beiden Dimensionen. Für die Bildungsgänge zeigte sich die charakteristische Leistungshierarchie mit den höchsten Mittelwertdifferenzen jeweils in Relation zur Hauptschule (Referenzkategorie, in Tabelle 2.4 nicht

dargestellt) für das Gymnasium, gefolgt von der Realschule und der Integrierten Gesamtschule. Für die Dimension *Systeme analysieren und entwerfen* war der Leistungsunterschied zwischen Hauptschule und Gymnasium deskriptiv betrachtet stärker ausgeprägt im Vergleich zu den anderen Dimensionen. Allerdings überlappten sich die entsprechenden Konfidenzintervalle, sodass von keinen signifikanten Unterschieden ausgegangen werden konnte. Demgegenüber wurde bei der Dimension *Entscheidungen treffen* mit 82 Prozent ein höherer Anteil an der Varianz auf Schulebene durch die Bildungsgänge aufgeklärt als bei den beiden anderen Dimensionen. Betrachtet man die Ergebnisse für Familiensprache und Geschlecht (siehe Tabelle 2.4), so zeigen sich für alle Dimensionen Mittelwertunterschiede zugunsten der Schülerinnen und Schüler mit Deutsch als Familiensprache. Diese Unterschiede waren zwar für die Dimension *Systeme analysieren und entwerfen* deskriptiv betrachtet geringer ausgeprägt, allerdings zeigte ein Vergleich der entsprechenden Konfidenzintervalle keine signifikanten Unterschiede. Für das Geschlecht zeigte sich auf der Dimension *Fehler suchen* ein signifikanter Mittelwertunterschied zuungunsten der Mädchen ( $\Delta M = -0.149$ ,  $KI_{95} = -0.230/-0.069$ ), während sich für die beiden anderen Dimensionen nur tendenzielle Geschlechtsunterschiede zeigen ließen, die statistisch nicht signifikant waren. Ein Vergleich der entsprechenden Konfidenzintervalle zeigte jedoch auch hier keine signifikanten Unterschiede zwischen den Dimensionen. Für die fachbezogenen Kompetenzen in der Mathematik, den Naturwissenschaften und im Lesen sowie für die fluiden Fähigkeiten (siehe Tabelle 2.4) zeigten sich für alle Dimensionen substantielle Korrelationen mit den Personenparametern, die deskriptiv für die Dimension *Systeme analysieren und entwerfen* durchweg höher ausfielen als für die Dimension *Entscheidungen treffen* und für diese wiederum höher als für die Dimension *Fehler suchen*. Ein Vergleich der Konfidenzintervalle zeigte jedoch nur zwischen den Dimensionen *Systeme analysieren und entwerfen* und *Fehler suchen* signifikante Unterschiede.

Die in Tabelle 2.4 präsentierten Ergebnisse liefern Hinweise auf differenzielle Zusammenhänge der Dimensionen des dreidimensionalen Modells mit externen Variablen, wenngleich sich diese nicht in allen Fällen inferenzstatistisch absichern lassen. Hypothese 2 konnte damit zumindest teilweise bestätigt werden.

**Tabelle 2.4: Intraklassenkorrelationen (ICC) und Mittelwertdifferenzen der Personenparameter des dreidimensionalen Modells der analytischen Problemlösekompetenz nach Bildungsgang, Familiensprache und Geschlecht sowie Korrelationen mit fachbezogenen Kompetenzen und fluiden Fähigkeiten (95-%-Konfidenzintervalle in eckigen Klammern)**

	Entscheidungen treffen	Systeme analysieren und entwerfen	Fehler suchen
ICC	.24	.32	.22
Bildungsgang			
Gesamtschule	0.341[0.168/0.514]***	0.379[0.156/0.602]**	0.405[0.193/0.616]***
Realschule	0.580[0.446/0.714]***	0.645[0.492/0.798]***	0.576[0.438/0.714]***
Gymnasium	1.163[1.046/1.280]***	1.293[1.156/1.429]***	1.067[0.949/1.184]***
$R^2$	.82	.76	.74
Familiensprache (1 = Deutsch)	0.532[0.357/0.707]***	0.410[0.255/0.566]***	0.558[0.406/0.710]***
Geschlecht (1 = weiblich)	-0.045[-0.117/0.027]	-0.024[-0.090/0.042]	-0.149[-0.230/-0.069]***
Mathematik	.611[.578/.644]***	.655[.624/.687]***	.568[.536/.600]***
Naturwissenschaften	.582[.551/.612]***	.621[.584/.657]***	.530[.494/.567]***
Lesen	.576[.544/.608]***	.608[.570/.646]***	.504[.469/.540]***
Fluide Fähigkeiten	.464[.424/.503]***	.481[.439/.523]***	.394[.354/.434]***

*Anmerkungen.* Geschätzte Mittelwertdifferenzen der z-standardisierten WL-Parameter für Bildungsgang (Referenzkategorie: Hauptschule), Familiensprache und Geschlecht sowie Korrelationen mit Mathematik, Naturwissenschaften, Lesen und fluide Fähigkeiten (für die Domänen Mathematik, Naturwissenschaften, Lesen und fluide Fähigkeiten wurden die WL-Parameter mit jedem der fünf Plausible Values pro Domäne aus PISA 2003 [OECD, 2005] korreliert und diese Korrelationen im Anschluss für jede Domäne gemittelt); \*\*\* $p < .001$ , \*\* $p < .01$ ;  $N = 2343$ .

## 2.6 Diskussion

### 2.6.1 Zusammenfassung der Ergebnisse

Gegenstand von Studie 1 war die Untersuchung von Aspekten der faktoriellen Validität der Modellierung analytischer Problemlösekompetenz. Hierzu wurde zum einen geprüft, inwiefern die im Rahmen von PISA 2003 eingesetzten drei Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* empirisch voneinander trennbare Dimensionen abbilden. Zum anderen wurde untersucht, ob sich Hinweise auf differenzielle Zusammenhänge zwischen diesen Dimensionen und externen Variablen finden lassen. Hypothese 1, nach der ein dreidimensionales Modell entsprechend den drei Arten von Problemstellungen die Datenstruktur adäquater abbilden sollte als ein eindimensionales Modell, konnte dabei bestätigt werden. Hypothese 2, der zufolge sich für die Dimensionen des dreidimensionalen Modells differenzielle Zusammenhänge mit externen Variablen zeigen sollten, konnte teilweise bestätigt werden.

### 2.6.2 Theoretische und praktische Implikationen

Die Ergebnisse zum Itemfit sowie zu den Reliabilitäten und Varianzen der latenten Dimensionen zeigen sowohl für das eindimensionale als auch für das dreidimensionale Modell zufriedenstellende Werte, sodass keines der Modelle eindeutig bevorzugt werden kann. Die Reliabilitäten des dreidimensionalen Modells liegen zwar etwas unter der Reliabilität des eindimensionalen Modells, allerdings erscheint dieser Unterschied insbesondere vor dem Hintergrund der geringen Itemanzahl pro Dimension im dreidimensionalen Modell zu gering, als dass er prinzipiell gegen dieses Modell sprechen würde. Wie in Abschnitt 2.2 dargestellt, bestehen neben deutlichen Unterschieden auch große Überlappungen zwischen den verschiedenen Arten von Problemstellungen hinsichtlich der zu erreichenden Ziele, der beteiligten Prozesse und der potenziell schwierigkeiterzeugenden Merkmale (vgl. Leutner et al., 2004). So ist es nicht verwunderlich, dass die Dimensionen des dreidimensionalen Modells hoch korreliert sind. Eine besondere Nähe zeigt sich hierbei für die Dimensionen *Entscheidungen treffen* und *Systeme analysieren und entwerfen*. Dies mag

unter anderem auf die ähnliche Grundkonzeption dieser beiden Arten von Problemstellungen zurückzuführen sein. Denn es handelt sich, wie in Abschnitt 2.2 dargestellt, in beiden Fällen um Projektaufgaben, ähnlich denjenigen die bereits in PISA 2000 eingesetzt wurden, jedoch mit dem Unterschied, dass jedes Projekt hier auf nur wenige Items beschränkt wird. Allerdings bewegen sich die Korrelationen in einer ähnlichen Größenordnung, wie man sie auch bei Kompetenzen aus unterschiedlichen fachlichen Domänen findet (z. B. OECD, 2005, 2012), sodass sie nicht per se gegen eine empirische Trennbarkeit der drei Arten von Problemstellungen sprechen. Das Ergebnis des LQT sowie die Werte des AIC und BIC favorisieren hingegen das dreidimensionale Modell, welches in der Gesamtschau als das Modell gelten kann, welches die Binnenstruktur der analytischen Problemlösekompetenz am adäquatesten abbildet (für Befunde zur Dimensionalität dynamischer bzw. komplexer Problemlösekompetenz siehe z. B. Greiff, Wüstenberg & Funke, 2012; Kröner, Plass & Leutner, 2005).

Welche praktischen Implikationen würde nun eine Kompetenzmessung zur Folge haben, die auf einer empirischen Trennung der drei Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* basierend? Die in Tabelle 2.4 dargestellten Ergebnisse liefern Hinweise auf differenzielle Zusammenhänge mit externen Variablen, auch wenn sich diese nicht für alle untersuchten Variablen inferenzstatistisch absichern lassen. Ein Grund hierfür ist sicherlich, wie bereits angesprochen, die relativ geringe Itemanzahl pro Dimension und die daraus resultierenden etwas verringerten Reliabilitäten gegenüber dem eindimensionalen Modell. Vor allem der Geschlechtsunterschied zuungunsten der Mädchen auf der Dimension *Fehler suchen* sowie die geringere Nähe zu fachbezogenen Kompetenzen und fluiden Fähigkeiten, insbesondere im Vergleich zur Dimension *Systeme analysieren und entwerfen*, fällt hierbei auf. Diese Ergebnisse lassen vermuten, dass eine diagnostische Profilbildung auf der Basis von drei getrennten Dimensionen entsprechend den drei Arten von Problemstellungen differenziertere Informationen liefern könnte als bei Verwendung aggregierter Informationen auf Basis einer gemeinsamen Dimension. Dies könnte insbesondere im Rahmen individualdiagnostischer Fragestellungen, bei denen spezifische Stärken und Schwächen im Fokus stehen, oder auch im Hinblick auf formative und summative Assessments auf Schul- und Unterrichtsebene relevant sein (vgl. Bögeholz & Eggert, 2013; Klieme & Leutner, 2006; Koeppen et al.,



2008). Gleiches gilt, wenn es um die Frage geht, mit welchen didaktischen Maßnahmen über eine Verbesserung der fächerübergreifenden analytischen Problemlösekompetenz auch ein Transfer auf fachbezogene Kompetenzen zum Beispiel in der Mathematik erreicht werden kann. Die unterschiedliche Nähe der einzelnen Dimensionen der analytischen Problemlösekompetenz zu fachbezogenen Kompetenzen lässt vermuten, dass Fördermaßnahmen auch unterschiedlich wirksam sein könnten, je nachdem auf welche Dimension sie fokussieren. Andererseits sind die berichteten Unterschiede hinsichtlich der Zusammenhänge mit externen Variablen keineswegs derart deutlich, als dass eine Trennung der drei Arten von Problemstellungen bei der Messung der analytischen Problemlösekompetenz zwingend erscheinen würde. Dies gilt insbesondere dann, wenn es wie bei internationalen Schulleistungsstudien häufig um die Bereitstellung von Informationen auf Systemebene geht (vgl. Leutner et al., 2007; Pant, 2013). In diesem Zusammenhang erscheint der Informationsverlust, verbunden mit der Modellierung der analytischen Problemlösekompetenz als eindimensionales Konstrukt, vernachlässigbar zu sein.

Zusammenfassend lässt sich festhalten, dass analytische Problemlösekompetenz je nach Forschungskontext und intendierter Verwendung der Testergebnisse entweder als dreidimensionales oder als eindimensionales Konstrukt modelliert werden kann. Steht eine differenziertere Betrachtung von individuellen Stärken und Schwächen beispielsweise im Rahmen individualdiagnostischer Fragestellungen im Vordergrund, so erscheint eine dreidimensionale Modellierung entsprechend den drei Arten von Problemstellungen angebracht. Richtet sich das Augenmerk hingegen auf die Bereitstellung von aggregierten Informationen, beispielsweise im Rahmen von Schulleistungsstudien, scheint eine eindimensionale Modellierung hinreichend zu sein.

### 2.6.3 Grenzen und Ausblick

Durch das sequenzielle Vorgehen bei der Untersuchung von Hypothese 2, bei dem zunächst ein Messmodell spezifiziert und im Anschluss daran die Zusammenhänge der geschätzten Personenparameter mit externen Variablen sowie Mittelwertdifferenzen für verschiedene Gruppen bestimmt wurden (für ein ähnliches Vorgehen siehe z. B. Hartig & Höhler, 2008; vgl. Sonnleitner, Keller, Martin & Brunner, 2013), kann die geclusterte Struktur

der verwendeten Daten berücksichtigt werden. Es erlaubt jedoch keine messfehlerfreie Modellierung der Korrelationen und Mittelwerte auf latenter Ebene. Diese hätte die aufwendige Spezifizierung von Messmodellen auch für die Tests zur Erfassung der fachbezogenen Kompetenzen und der fluiden Fähigkeiten und damit eine komplette Neuskalierung der verwendeten PISA 2003 Daten erforderlich gemacht. Auf ein solches Vorgehen wurde daher verzichtet, zumal mit den Plausible Values aus PISA 2003 bereits methodisch elaborierte Maße für die betreffenden fachbezogenen Kompetenzskalen vorlagen.

Eine weitere Einschränkung von Studie 1 ergibt sich aus der relativ geringen Itemanzahl des PISA-2003-Problemlösetests. So zeigen die präsentierten Ergebnisse, dass eine getrennte Modellierung der drei Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* zwar möglich ist, allerdings eignet sich dieser Test nicht zur Untersuchung individualdiagnostischer Fragestellungen auf Basis einer entsprechenden dreidimensionalen Modellierung. Auch wenn die Reliabilitäten des dreidimensionalen Modells für die durchgeführten Dimensionalitätsanalysen zur Untersuchung der faktoriellen Validität hinreichend hoch sind, ist für die adäquate Operationalisierung der drei Dimensionen im Rahmen der Kompetenzdiagnostik als Spezialfall der Individualdiagnostik ein größerer Itempool notwendig. Nur auf diese Weise können die hierfür notwendigen höheren Reliabilitätswerte der einzelnen Dimensionen gewährleistet werden (vgl. Leutner et al., 2007).

Mit einem entsprechend erweiterten Itempool könnte dann in zukünftigen Untersuchungen dezidiert eruiert werden, welche zusätzlichen Informationen eine Modellierung von drei getrennten Dimensionen nach den drei Arten von Problemstellungen für die Kompetenzdiagnostik tatsächlich liefert. Für eine möglichst umfassende Untersuchung der faktoriellen Validität der Modellierung analytischer Problemlösekompetenz wäre es in zukünftigen Untersuchungen zudem empfehlenswert, weitere Arten von Problemstellungen in Dimensionalitätsanalysen zu berücksichtigen. Hierzu könnten neben den in PISA 2003 verwendeten *fächerverbindenden* Problemstellungen beispielsweise auch *fächerunabhängige* Problemstellungen, wie sie in der nationalen Ergänzungsstudie zu PISA 2000 eingesetzt wurden (siehe Abschnitt 2.2) zur Anwendung kommen. So ließe sich prüfen, inwiefern diese beiden Arten von Problemstellungen tatsächlich dasselbe Konstrukt abbilden und wie viele

Dimensionen zur adäquaten Modellierung der analytischen Problemlösekompetenz notwendig sind.

## 2.7 Literatur

- Adams, R. J. & Khoo, S. T. (1996). *Quest: The interactive test analysis system*. Melbourne, Australia: Australian Council for Educational Research.
- Adams, R. J., Wilson, M. R. & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1–23. doi: 10.1177/0146621697211001
- Bayrhuber, M., Leuders, T., Bruder, R. & Wirtz, M. (2010). Repräsentationswechsel beim Umgang mit Funktionen - Identifikation von Kompetenzprofilen auf der Basis eines Kompetenzstrukturmodells. Projekt HEUREKO. *Zeitschrift für Pädagogik*, 56. Beiheft, 28–39.
- Bögeholz, S. & Eggert, S. (2013). Welche Rolle spielt Kompetenzdiagnostik im Rahmen von Lehr-Lernprozessen? *Zeitschrift für Erziehungswissenschaft*, 16 (suppl. 1), 59–64. doi: 10.1007/s11618-013-0384-2
- Carstensen, C. H., Knoll, S., Rost, J. & Prenzel, M. (2004). Technische Grundlagen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 371–387). Münster: Waxmann.
- Ebach, J., Klieme, E. & Hensgen, A. (1999). *Der SL HAM 6/7 Problemlösetest*. Bonn: Institut für Bildungsforschung.
- Fleischer, J., Koeppen, K., Kenk, M., Klieme, E. & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. *Zeitschrift für Erziehungswissenschaft*, 16 (suppl. 1), 5–22. doi: 10.1007/s11618 013-0379-z
- Fleischer, J., Wirth, J., Rumann, S. & Leutner, D. (2010). Strukturen fächerübergreifender und fachlicher Problemlösekompetenz – Analyse von Aufgabenprofilen. *Zeitschrift für Pädagogik*, 56. Beiheft, 239–248.
- Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester, UK: Wiley.

- Greiff, S., Wüstenberg, S. & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, 36, 189–213. doi: 10.1177/0146621612439620
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Berlin: Springer. doi: 10.1007/978-3-642-20072-4\_7
- Hartig, J. & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. *Zeitschrift für Psychologie / Journal of Psychology*, 216, 89–101. doi: 10.1027/0044-3409.216.2.89
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer. doi: 10.1007/3-540-33020-8\_9
- Heller, K. A. & Perleth, C. (2000). *KFT 4-12+R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Hogrefe.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan & J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.
- Hox, J. J. (2002). *Multilevel Analysis*. Mahwah, NJ: Erlbaum.
- Jude, N., Klieme, E., Eichler, W., Lehmann, R. H., Nold, G., Schröder, S. et al. (2008). Strukturen sprachlicher Kompetenzen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 191–201). Weinheim: Beltz.
- Klieme, E., Hartig, J. & Wirth, J. (2005). Analytisches Problemlösen: Messansatz und Befunde zu Planungs- und Entscheidungsaufgaben. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 37–54). Wiesbaden: VS.
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52, 876–903.

- Klieme, E., Leutner, D. & Kenk, M. (Hrsg.). (2010). Kompetenzmodellierung – Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes [Themenheft]. *Zeitschrift für Pädagogik*, 56. Beiheft.
- Koeppen, K., Hartig, J., Klieme, E. & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie / Journal of Psychology*, 216, 61–73. doi: 10.1027/0044-3409.216.2.61
- Kröner, S., Plass, J. L. & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, 33, 347–368. doi: 10.1016/j.intell.2005.03.002
- Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods and Research*, 33, 188–229. doi: 10.1177/0049124103262065
- Lehmann, R. H., Peek, R., Gänsfuß, R. & Husfeldt, V. (2002). *Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 9. Ergebnisse einer Längsschnittuntersuchung in Hamburg*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Leutner, D., Fleischer, J., Spoden, C. & Wirth, J. (2007). Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik. *Zeitschrift für Erziehungswissenschaft, Sonderheft 8*, 149–167. doi: 10.1007/978-3-531-90865-6\_9
- Leutner, D., Klieme, E., Fleischer, J. & Kuper, H. (Hrsg.). (2013). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Aktuelle Diskurse im DFG-Schwerpunktprogramm. *Zeitschrift für Erziehungswissenschaft*, 16 (suppl. 1).
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147–175). Münster: Waxmann.
- Leutner, D., Wirth, J., Klieme, E. & Funke, J. (2005). Ansätze zur Operationalisierung und deren Erprobung im Feldtest zu PISA 2000. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 21–36). Wiesbaden: VS.

- Masters, G. N. (1982). A Rasch Model for partial credit scoring. *Psychometrika*, 47, 149–174. doi: 10.1007/BF02296272
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. doi: 10.1037/0003-066X.50.9.741
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29, 133–161. doi: 10.1111/j.1745-3984.1992.tb00371.x
- Neumann, K. (2013). Mit welchem Auflösungsgrad können Kompetenzen modelliert werden? In welcher Beziehung stehen Modelle zueinander, die Kompetenz in einer Domäne mit unterschiedlichem Auflösungsgrad beschreiben? *Zeitschrift für Erziehungswissenschaft*, 16 (suppl. 1), 35–39. doi: 10.1007/s11618-013-0382-4
- OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing.
- OECD. (2004). *Problem solving for tomorrow's world: First measurements of cross-curricular competencies from PISA 2003*. Paris: OECD Publishing.
- OECD. (2005). *PISA 2003: Technical report*. Paris: OECD Publishing.
- OECD. (2012). *PISA 2009: Technical report*. Paris: OECD Publishing.
- Pant, H. A. (2013). Wer hat einen Nutzen von Kompetenzmodellen? *Zeitschrift für Erziehungswissenschaft*, 16 (suppl. 1), 71–79. doi: 10.1007/s11618-013-0388-y
- PISA-Konsortium Deutschland (Hrsg.). (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Prenzel, M., Drechsel, B., Carstensen, C. H. & Ramm, G. (2004). PISA 2003 – eine Einführung. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 13–46). Münster: Waxmann.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: The Danish Institute for Educational Research.

- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T. & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Chicago, IL: SSI.
- Rost, D. H. (2009). *Intelligenz Fakten und Mythen*. Weinheim: Beltz.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (2. überarb. und erw. Aufl.). Bern: Huber.
- Schnotz, W., Horz, H., McElvany, N., Schroeder, S., Ullrich, M., Baumert, J. et al. (2010). Das BITE-Projekt: Integrative Verarbeitung von Bildern und Texten in der Sekundarstufe I. *Zeitschrift für Pädagogik*, 56. Beiheft, 143–153.
- Sonnleitner, P., Keller, U., Martin, R. & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41, 289–305. doi: 10.1016/j.intell.2013.05.002
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450. doi: 10.1007/BF02294627
- Wilson, M. R. (2005). *Constructing measures. An item response modeling approach*. Mahwah, NJ: Erlbaum.
- Winkelmann, H., Robitzsch, A., Stanat, P. & Köller, O. (2012). Mathematische Kompetenzen in der Grundschule. Struktur, Validierung und Zusammenspiel mit allgemeinen kognitiven Fähigkeiten. *Diagnostica*, 58, 15–30. doi: 10.1026/0012-1924/a000061
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Wu, M. L. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31, 114–128. doi: 10.1016/j.stueduc.2005.05.005
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell, Australia: ACER Press.



## 3 STUDIE 2: DISKRIMINANTE VALIDITÄT DER MODELLIERUNG ANALYTISCHER PROBLEMLÖSEKOMPETENZ – ABGRENZUNG VON FLUIDEN FÄHIGKEITEN

### 3.1 Einleitung

Neben der Frage nach der faktoriellen Validität beziehungsweise Dimensionalität ist die Abgrenzung von anderen Konstrukten, also die Frage nach der *diskriminanten Validität* (Campbell & Fiske, 1959), ein weiterer zentraler Aspekt der Konstruktvalidität der Modellierung von Kompetenzen (*externer Aspekt* der Konstruktvalidität, sensu Messick, 1995). Im Falle der analytischen Problemlösekompetenz steht dabei insbesondere die fundierte theoretische und empirische Verortung im Vergleich zu allgemeinen kognitiven Grundfähigkeiten, also von Intelligenzmaßen im Fokus. Zentral ist hierbei insbesondere die Frage nach der Abgrenzbarkeit von fluiden Fähigkeiten beziehungsweise schlussfolgerndem Denken (*reasoning*) als bewährte und erklärungskräftige Konstrukte mit gleichzeitig deutlicher Nähe zur Problemlösekompetenz. Dies zeigt auch die Debatte zur diskriminanten Validität der Modellierung fachbezogener und fächerübergreifender Kompetenzen im Verhältnis zu Intelligenzmaßen vor dem Hintergrund international vergleichender Schulleistungsstudien (vgl. Baumert, Brunner, Lüdtke & Trautwein, 2007; Baumert, Lüdtke, Trautwein & Brunner, 2009; Gottfredson, 2003; Rindermann, 2006; Prenzel, Walter & Frey, 2007).

Gegenstand von Studie 2 ist demnach die Untersuchung von Aspekten der diskriminanten Validität der Modellierung analytischer Problemlösekompetenz gegenüber fluiden Fähigkeiten. Hierzu werden konkurrierende Strukturmodelle bezüglich des Zusammenhangs von analytischer Problemlösekompetenz und fluiden Fähigkeiten verglichen sowie Zusammenhänge mit externen Variablen untersucht.

## 3.2 Theoretischer Hintergrund

### 3.2.1 Allgemeine kognitive Grundfähigkeiten – Intelligenz

Für das Konstrukt der Intelligenz existiert eine Fülle unterschiedlicher Konzeptualisierungen, die sich mitunter beachtlich voneinander unterscheiden und somit keine allgemein anerkannte beziehungsweise einheitliche Definition von Intelligenz zulassen (für eine Übersicht siehe z. B. Rost, 2009, Kap. 1). Dennoch besteht weitestgehend Konsens darüber, dass Intelligenz eine allgemeine und sehr breite kognitive Fähigkeit beschreibt, die unter anderem die Fähigkeit zum schlussfolgernden und abstrakten Denken, zum Planen, zum Verstehen komplexer Ideen, zur Aneignung von Wissen sowie zum Problemlösen umfasst (Eysenck, 2004; Snyderman & Rothman, 1987; siehe auch Rost, 2009, Kap. 1). Kognitive Fähigkeitskonstrukte wie die Intelligenz werden klassischerweise mithilfe faktorieller Strukturmodelle beschrieben (z. B. Cattell, 1963; Deary, 2012; Horn, 1994). Spearman (1927) postulierte eine allgemeine kognitive Fähigkeit *g*, die die Zusammenhänge zwischen verschiedenen Aufgaben und den zu deren Bewältigung notwendigen intellektuellen Fähigkeiten, erklärt. Cattell (1963) und – darauf aufbauend – Horn (1994) gingen hingegen von zwei Intelligenzfacetten, den *fluiden* und *kristallinen* Fähigkeiten, aus. Fluide Fähigkeiten bezeichnen dabei die Fähigkeiten zum schlussfolgernden Denken bei der Bearbeitung neuartiger Aufgaben und Problemstellungen. Kristalline Fähigkeiten bezeichnen demgegenüber die durch Lernen und Prozesse der Sozialisation erworbenen Fähigkeiten und Wissensbestände. Fluide und kristalline Fähigkeiten wurden in der Folge noch um weitere Intelligenzfacetten ergänzt (z. B. Horn & Masunaga, 2006; Horn & Noll, 1997). Eine Zusammenfassung dieser Forschungsansätze und gleichzeitig eine der umfassendsten Taxonomien kognitiver Fähigkeiten stellt die Drei-Stratum-Theorie von Carroll (1993) dar, die auf einer Reanalyse von über 400 Datensätzen basiert. Carroll zufolge lassen sich drei Generalisierungsebenen kognitiver Fähigkeiten (*strata*) unterscheiden. Auf der obersten Ebene befindet sich die allgemeine kognitive Fähigkeit, die komplexe kognitive Prozesse wie Aufmerksamkeitssteuerung, Planung oder Arbeitsgedächtniskapazität beschreibt (Carroll, 1993). Auf der mittleren Ebene werden acht breite kognitive Sekundärfähigkeiten unterschieden (u. a. fluide und kristalline Fähigkeiten), und auf der untersten Ebene identifiziert Carroll eine große Anzahl kognitiver Primärfähigkeiten. Fluide Fähigkeiten gelten gemeinhin als guter Indikator für

den *g*-Faktor (sensu Spearman) beziehungsweise für allgemeine kognitive Fähigkeiten (sensu Carroll) oder werden gar mit diesen gleichgesetzt (Carlstedt, 2001; Gustafsson, 1988). Zentraler Primärfaktor der fluiden Fähigkeiten ist das schlussfolgernde Denken (*reasoning*), welches häufig als Kern der Intelligenz betrachtet wird (z. B. Carroll, 1993; Mackintosh, 1998) und beispielsweise mit figuralen Matrizenaufgaben oder figuralen Analogieaufgaben (*figural reasoning*) erfasst wird (z. B. Heller & Perleth, 2000; Liepmann, Beauducel, Brocke & Amthauer, 2007; siehe auch Gustafsson, 1988; Brunner, Krauss & Kunter, 2008). Modelle, die ähnlich wie Carroll von einem hierarchischen Aufbau kognitiver Fähigkeiten ausgehen, sind weitestgehend akzeptiert und stellen derzeit den *state of the art* in der Intelligenzforschung dar (z. B. Deary, 2012; McGrew, 2009; Schulze, 2005).

### 3.2.2 Problemlösen und allgemeine kognitive Grundfähigkeiten

Das Konzept der Intelligenz, insbesondere der fluiden Fähigkeiten, lässt eine deutliche Nähe zur analytischen Problemlösekompetenz erkennen. So werden beim Problemlösen hohe Anforderungen an das analytische, schlussfolgernde und kombinatorische Denken sowie an das analoge Schließen gestellt (OECD, 2003, 2004), die ebenfalls für die Intelligenz zentral sind, insbesondere für die figuralen Fähigkeiten (vgl. Rost, 2009, Kap. 2; Wilhelm & Engle, 2005). Daher verwundert es nicht, dass im Rahmen der Intelligenzforschung Problemlösen häufig als kognitive Aktivität beschrieben wird, die eng mit fluiden Fähigkeiten assoziiert ist (Kyllonen & Lee, 2005; Rost, 2009, Kap. 1; vgl. Carroll, 1993). Fasst man Intelligenz, so wie dies Sternberg (1985, 2012) in seiner triarchischen Intelligenztheorie tut, sehr viel breiter auf als dies in der Forschung üblicherweise geschieht, so werden Problemlösen und Intelligenz sogar zu praktisch deckungsgleichen Konstrukten (Klieme, Funke, Leutner, Reimann & Wirth, 2001).

Allerdings lässt sich analytisches Problemlösen im Sinne einer Kompetenz konzeptuell deutlich von fluiden Fähigkeiten trennen:

(1) *Bereichs- oder Situationsspezifität*: Während fluide Fähigkeiten mit einer sehr hohen Generalisierbarkeit assoziiert werden (z. B. Ree & Carretta, 2002), ist die analytische

Problemlösekompetenz, wie alle Kompetenzen stärker auf bestimmte Kontexte und Anforderungsbereiche bezogen<sup>3</sup> (Leutner, Funke, Klieme & Wirth, 2005).

(2) *Prinzipielle Erlernbarkeit*: Intelligenz, insbesondere im Sinne der fluiden Fähigkeiten, gilt als zeitlich relativ stabiles Konstrukt (Rost, 2009, Kap. 8). Ein zentrales Merkmal der analytischen Problemlösekompetenz ist hingegen ihre Veränderbarkeit und prinzipielle Erlernbarkeit (Klieme et al., 2001; Leutner, Funke et al., 2005; siehe auch Buchwald, Fleischer, Rumann, Wirth & Leutner, 2017).

(3) *Binnenstruktur an Anforderungen orientiert*: Bei der Untersuchung der Binnenstruktur und Dimensionalität fluider Fähigkeiten und der Intelligenz insgesamt werden häufig grundlegende kognitive und biologische Prozesse betrachtet (Wilhelm & Engle, 2005). Die Dimensionalität der analytischen Problemlösekompetenz wie von Kompetenzen allgemein wird hingegen stärker durch die in bestimmten Situationen zu bewältigenden Anforderungen definiert (Hartig & Klieme, 2006; vgl. Fleischer, Koeppen, Kenk, Klieme & Leutner, 2013), wie beispielsweise anhand der in Studie 1 beschriebenen, unterschiedlichen Arten von Problemstellungen deutlich wird.

Auch ein Vergleich der Aufgaben aus Testinstrumenten zur Erfassung fluider Fähigkeiten mit Aufgaben zur Erfassung der analytischen Problemlösekompetenz (siehe Abschnitt 1.3.3) zeigt deutliche Unterschiede: So sind Testaufgaben zur Erfassung fluider Fähigkeiten, wie beispielsweise figurale Matrizenaufgaben oder figurale Analogieaufgaben (z. B. Heller & Perleth, 2000; Liepmann et al., 2007), für gewöhnlich dekontextualisiert und weniger komplex als Problemlöseaufgaben, zu deren Lösung außerdem ein Mindestmaß an inhaltspezifischem Wissen notwendig ist (Klieme, 2004; Leutner, Funke et al., 2005; Wilhelm, 2005).

---

<sup>3</sup> Einschränkung sei darauf hingewiesen, dass dieser prinzipielle Unterschied zwischen Kompetenzen und fluiden Fähigkeiten im Falle der analytischen Problemlösekompetenz als fächerübergreifendem Konstrukt deutlich abgeschwächt ist. Dennoch ist für analytische Problemlösekompetenz, wie sie in PISA 2003 operationalisiert wurde, in Bereichen außerhalb des Schulkontexts nicht zwangsläufig von prognostischer Validität, zumindest nicht in dem Maße, wie es für fluide Fähigkeiten, z. B. in Bezug auf Studienerfolg und beruflichen Erfolg, vielfach gezeigt werden konnte (für einen Überblick siehe Rost, 2009, Kap. 6). Insofern kann trotz der hohen Generalisierbarkeit beider Konstrukte eine stärkere Bereichs- oder Situationsspezifität der analytischen Problemlösekompetenz angenommen werden (siehe auch Klieme et al., 2001).

Die beschriebenen konzeptuellen Ähnlichkeiten und Unterschiede zwischen Intelligenz, insbesondere den fluiden Fähigkeiten, und der analytischen Problemlösekompetenz schlagen sich auch empirisch nieder. So ist die Befundlage zum korrelativen Zusammenhang zwischen Intelligenzmaßen und den Leistungen in traditionellen (einfachen) Problemlöseaufgaben heterogen (für eine Übersicht siehe z. B. Süß, 1996). Als Grund dafür kann, neben unterschiedlichen Operationalisierungen von Problemlösen und Intelligenz, ein umgekehrt u-förmiger Verlauf des Zusammenhangs angenommen werden (Raaheim, 1988). Demnach ist die Korrelation zwischen Intelligenz und Problemlösen bei mittlerer Vertrautheit mit der Problemstellung am höchsten, während bei völlig neuartigen Problemstellungen, die per Versuch und Irrtum gelöst werden, sowie bei bereits bekannten Problemstellungen (Routineaufgaben), die allein mit vorhandenem Wissen gelöst werden, kaum Intelligenz notwendig ist und die Korrelation damit wesentlich geringer ausfällt (Leutner, 2002; Süß, 1996). Im Hinblick auf den Zusammenhang von fluiden Fähigkeiten und analytischer Problemlösekompetenz konnten Leutner, Wirth, Klieme und Funke (2005) anhand der Feldtestdaten aus PISA 2000 zeigen, dass sich die Leistungen bei Aufgaben dynamischen Problemlösens, analytischen Problemlösens und fluider Fähigkeiten, trotz hoher Korrelationen, nur sehr schlecht durch einen gemeinsamen latenten Faktor erklären lassen. Ein Modell mit drei latenten Dimensionen bildet die Datenstruktur dagegen wesentlich besser ab (siehe auch die Ergebnisse zur Trennbarkeit von fluiden Fähigkeiten, Problemlösekompetenz und fachbezogener Kompetenzen mithilfe multidimensionaler Skalierungen; Leutner, Klieme, Meyer & Wirth, 2004; Wirth, Leutner & Klieme, 2005). Die mitunter hohen Korrelationen von analytischer Problemlösekompetenz und fluiden Fähigkeiten sind in der Regel jedoch deutlich von 1 verschieden und fallen zudem auch nicht höher aus, als man es für Korrelationen zwischen Kompetenzen unterschiedlicher fachlicher Domänen erwarten würde (z. B. für PISA 2003: Mathematik und Lesen,  $r = .77$ ; Mathematik und Naturwissenschaften,  $r = .82$ ; Lesen und Naturwissenschaften,  $r = .83$ ; OECD, 2005, Kap. 13). So fanden Wirth et al. (2005) anhand einer Substichprobe aus PISA 2000 eine Korrelation von  $r = .61$  auf manifester Ebene, Leutner et al. (2004) anhand einer Substichprobe aus PISA 2003 eine Korrelation von  $r = .72$  auf Ebene der latenten Konstrukte und Scherer und Tiemann (2014) bei Schülerinnen und Schülern der 8. und 10. Jahrgangsstufe eine Korrelation von  $r = .60$ , ebenfalls auf manifester Ebene. Fischer, Greiff, Wüstenberg, Fleischer, Buchwald und

Funke (2015) wiesen bei 15-jährigen Schülerinnen und Schülern eine Korrelation von  $r = .56$  und bei Studierenden unterschiedlicher Fachrichtungen eine Korrelation von  $r = .40$  jeweils auf latenter Ebene nach. Neben korrelativen Zusammenhängen von analytischer Problemlösekompetenz und fluiden Fähigkeiten geben auch Vergleiche der jeweiligen Zusammenhänge mit externen Variablen Aufschluss über die Ähnlichkeiten und Unterschiede bei den Konstrukten. So konnten Leutner et al. (2004) anhand einer Teilstichprobe aus PISA 2003 für analytische Problemlösekompetenz höhere Korrelationen mit Mathematik ( $r = .90$ ), Lesen ( $r = .81$ ) und Naturwissenschaften ( $r = .85$ ) zeigen, als für fluide Fähigkeiten mit Mathematik ( $r = .74$ ), Lesen ( $r = .64$ ) und Naturwissenschaften ( $r = .68$ ). Analytische Problemlösekompetenz scheint demnach eine größere Nähe zu fachbezogenen Kompetenzen im Schulkontext aufzuweisen, als dies für fluide Fähigkeiten der Fall ist.

Wie in diesem Abschnitt dargestellt, weisen die fluiden Fähigkeiten und die analytische Problemlösekompetenz einerseits konzeptuell und empirisch hinreichend deutliche Unterschiede auf, sodass von diskriminanter Validität der Modellierung analytischer Problemlösekompetenz gegenüber den fluiden Fähigkeiten ausgegangen werden kann. Auf der anderen Seite zeigen sich jedoch auch deutliche Überschneidungen und Gemeinsamkeiten, die darauf hindeuten, dass zur erfolgreichen Bearbeitung von analytischen Problemlöseaufgaben ebenfalls fluide Fähigkeiten in nicht unerheblichem Maße notwendig sind. Es kann somit von einer multiplen Bedingtheit analytischer Problemlöseleistungen sowohl durch die analytische Problemlösekompetenz als auch durch fluide Fähigkeiten ausgegangen werden. Fluide Fähigkeiten können damit quasi als Komponente der analytischen Problemlösekompetenz aufgefasst werden, die je nach Art der zu bearbeitenden Problemstellung einen unterschiedlich starken Einfluss auf die Qualität der Problemlösung ausüben kann (Leutner, Funke et al., 2005).

Den zuvor beschriebenen Untersuchungen zum Zusammenhang von fluiden Fähigkeiten und analytischer Problemlösekompetenz liegen mehrdimensionale, nicht hierarchische Konzeptionen zugrunde. Hierbei werden beide Konstrukte als distinkte Dimensionen modelliert, bei denen die Testitems als Indikatoren jeweils nur einer Dimension zugeordnet werden. Diese Einfachladungsstruktur der Items (*between-item multidimensionality*; Adams, Wilson & Wang, 1997) wird im Folgenden als Standardmodell bezeichnet, da es sich

hierbei um das nach wie vor gängigste Modell in der pädagogisch-psychologischen Forschung sowie in der Bildungsforschung handelt (Bollen & Lennox, 1991; vgl. OECD, 2005; 2012). Das Standardmodell lässt jedoch zum einen Überlegungen zum hierarchischen Aufbau kognitiver Fähigkeiten, wie sie sich aus den beschriebenen Modellen allgemeiner kognitiver Grundfähigkeiten (z. B. Carroll, 1993) ergeben, außer Acht. Zum anderen werden Annahmen zur multiplen Bedingtheit von Problemlöseleistungen, die sich aus den vergleichbaren Anforderungen beim analytischen Problemlösen und bei fluiden Fähigkeiten ergeben (z. B. Kyllonen & Lee, 2005; Leutner, Funke et al., 2005), nicht berücksichtigt.

### 3.3 Fragestellungen und Hypothesen

Eine Möglichkeit zur Berücksichtigung der zuvor beschriebenen Annahmen hinsichtlich des Zusammenspiels von analytischer Problemlösekompetenz und fluiden Fähigkeiten bietet sich im Rahmen von Nested-Faktor-Modellen (Gustafsson & Balke, 1993; siehe auch Brunner, 2008; Hornung, Brunner, Reuter & Martin, 2011). Hierbei wird, neben einem allgemeinen Faktor<sup>4</sup> für fluide Fähigkeiten, auf dem sowohl Problemlöseitems als auch Items zur Erfassung fluider Fähigkeiten laden, zusätzlich ein genesteter Faktor spezifiziert, auf dem nur die Problemlöseitems laden (siehe Abbildung 3.1). Dadurch ergibt sich eine multiple Bedingtheit der Leistungen in den Problemlöseitems sowohl durch einen Faktor, der fluiden Fähigkeiten repräsentiert, als auch durch einen Faktor, der die um die fluiden Fähigkeiten residualisierte analytische Problemlösekompetenz erfasst (*within-item multidimensionality*; Adams et al., 1997). Die Modellierung von analytischer Problemlösekompetenz und fluiden Fähigkeiten mithilfe eines Nested-Faktor-Modells ermöglicht somit eine genauere Untersuchung der diskriminanten Validität der Erfassung analytischer Problemlösekompetenz, indem die Gesamtvarianz auf zwei voneinander unabhängige Dimensionen aufgeteilt wird. Die Berücksichtigung der multiplen Bedingtheit von Itembeantwortungen

---

<sup>4</sup> Die Begriffe *Dimension* und *Faktor* können synonym verwendet werden. Entsprechend der unterschiedlichen Forschungstraditionen wird jedoch im Kontext der Item-Response-Theorie in der Regel von einer *Dimension* gesprochen, wenn es um die Abbildung eines latenten Konstrukts geht, während im Kontext von faktoranalytischen Modellen und Strukturgleichungsmodellen in der Regel der Begriff *Faktor* verwendet wird. Dieser Unterscheidung soll in der vorliegenden Arbeit entsprochen werden.

im Rahmen eines Nested-Faktor-Modells erfüllt darüber hinaus eine der zentralen Forderungen der Kompetenzmodellierung nach elaborierteren Modellen der Interaktion von Personen und Testitems, wie sie beispielsweise von Koeppen, Hartig, Klieme und Leutner (2008) erhoben wurden.

Vor dem Hintergrund der zuvor geschilderten theoretischen Überlegungen wird davon ausgegangen, dass ein Nested-Faktor-Modell, das von einer multiplen Bedingtheit von analytischen Problemlöseleistungen durch analytische Problemlösekompetenz und fluide Fähigkeiten ausgeht, eine bessere Modellpassung aufweist als das Standardmodell, welches analytische Problemlöseleistungen und Leistungen in Items zur Erfassung fluider Fähigkeiten auf jeweils unterschiedliche Faktoren zurückführt (Hypothese 1). Entsprechend der Annahme, dass die Modellierung analytische Problemlösekompetenz diskriminante Validität gegenüber fluiden Fähigkeiten besitzt, beide Konstrukte also nicht deckungsgleich sind, sondern fluide Fähigkeiten vielmehr als Komponente der analytischen Problemlösekompetenz betrachtet werden können, wird davon ausgegangen, dass ein um fluide Fähigkeiten bereinigter Faktor der analytischen Problemlösekompetenz signifikante Varianz und substantielle Faktorladungen aufweist und somit substantiell und inhaltlich interpretierbar ist (Hypothese 2). Darüber hinaus wird davon ausgegangen, dass sich für diesen Faktor im Vergleich zum entsprechenden Faktor im Standardmodell differenzielle Zusammenhänge mit externen Variablen zeigen und er somit eine praktische Bedeutsamkeit aufweist (Hypothese 3).

## **3.4 Methode**

### **3.4.1 Stichprobe und Datengrundlage**

Für die Untersuchung wurden wie bereits in Studie 1 die Daten der deutschen PISA-Haupterhebung 2003 zum Problemlösen ( $N = 2343$ ; Probanden mit Testheften, in denen entsprechende Itemcluster enthalten waren) ausgewählt. Für die genauere Beschreibung der Stichprobe sei daher auf Abschnitt 2.4.1 verwiesen (siehe auch Prenzel, Drechsel, Carstensen & Ramm, 2004).



### 3.4.2 Instrumente und Indikatoren

Zur Erfassung der Problemlösekompetenz wurden anhand der Daten des PISA-Problemlösetests zunächst für die drei Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* (siehe Studie 1) jeweils getrennte eindimensionale Partial-Credit-Modelle mit dem Programm ConQuest 2.0 (Wu, Adams, Wilson & Haldane, 2007) geschätzt. Die insgesamt 18 Testitems gingen hierbei entsprechend der in PISA verwendeten Kodierung entweder als dichotome (0 = *falsch*; 1 = *richtig*) oder polytome Items (0 = *falsch*; 1 = *teilweise richtig*; 2 = *vollständig richtig*) in die Skalierung ein. Im Anschluss wurden für diese drei geschätzten Dimensionen Personenparameter (WL-Schätzer, Warm, 1989) bestimmt. Zur Erfassung der fluiden Fähigkeiten wurden die Daten der Subskala *N2 Figurenanalogien des Kognitiven Fähigkeitstests* (KFT 5-12+R, Heller & Perleth, 2000) verwendet. Diese Subskala erfasst schlussfolgerndes Denken (*reasoning*), welches als guter Indikator für fluide Fähigkeiten  $fF$  und ebenso für allgemeine kognitive Fähigkeiten  $g$  gilt (Horn & Noll, 1997; Rost, 2009, Kap. 2). Die 25 dichotomen Items (0 = *falsch*; 1 = *richtig*) dieser Skala wurden nach der Odd-Even-Methode in zwei Hälften geteilt. Anschließend wurden für diese beiden Hälften jeweils getrennte eindimensionale Rasch-Modelle mit dem Programm ConQuest 2.0 (Wu et al., 2007) geschätzt. Für beide geschätzten Dimensionen wurden sodann Personenparameter (WL-Schätzer, Warm, 1989) bestimmt. Für die Personenparameter aller Dimensionen konnten signifikante Varianzen geschätzt werden sowie im Falle der beiden Dimensionen des KFT auch zufriedenstellende Reliabilitäten (siehe Tabelle 3.1). Für die drei Dimensionen des Problemlösetests sind die Reliabilitäten aufgrund der relativ geringen Itemanzahl pro Dimension jedoch deutlich niedriger.<sup>5</sup> Zur Erfassung der Mathematikkompetenz, der naturwissenschaftlichen Kompetenz und der Lesekompetenz wurden die Daten der entsprechenden PISA-Skalen herangezogen (siehe PISA-Konsortium Deutschland, 2006, Teil I, Kap. 1-3). Darüber hinaus wurden die in PISA erhobenen Informationen zu Familiensprache, Geschlecht sowie Bildungsgang in den Analysen verwendet (PISA-Konsortium Deutschland, 2006, Teil II, Kap. 1 & 2).

---

<sup>5</sup> Aufgrund der getrennten Skalierung der drei Dimensionen analytischen Problemlösens liegen die Reliabilitäten zudem unter jenen, die für das dreidimensionale Modell in Studie 2 (siehe Abschnitt 2.5) geschätzt wurden.

**Tabelle 3.1: Reliabilitäten (EAP-Schätzungen), Varianzen ( $\sigma^2$ ) und Infit-Werte der Dimensionen der analytischen Problemlösekompetenz und der fluiden Fähigkeiten**

	Reliabilität	$\sigma^2$ (SE)	Infit (weighted MNSQ)
Problemlösen			
Entscheidungen treffen (6 Items)	.54	1.42 (.116)***	0.98–1.01
Systeme analysieren und entwerfen (7 Items)	.58	1.08 (.082)***	0.98–1.06
Fehler suchen (5 Items)	.57	3.74 (.305)***	0.87–1.16
Fluide Fähigkeiten ( <i>reasoning</i> )			
KFT-1 (13 Items)	.79	2.32 (.101)***	0.88–1.14
KFT-2 (12 Items)	.76	2.07 (.092)***	0.81–1.26

Anmerkungen. \*\*\* $p < .001$ ;  $N = 2343$ .

### 3.4.3 Vorgehen und statistische Analysen

Zur Überprüfung der Hypothesen 1 und 2 wurden drei konfirmatorische Faktormodelle mit dem Programm *Mplus* (Version 7; Muthén & Muthén, 1998–2012) spezifiziert, in denen die z-standardisierten WL-Schätzer der Personenparameter als manifeste Variablen verwendet wurden. Dieses Vorgehen ist robuster gegenüber Verletzungen von Verteilungsannahmen und vereinfacht die Schätzung der Modelle im Vergleich zur Verwendung von einzelnen Items als Indikatoren auf manifester Ebene (Little, Cunningham, Shahar & Widaman, 2002; vgl. Sonnleitner, Keller, Martin & Brunner, 2013). Vor dem Hintergrund der Ergebnisse von Studie 1 der vorliegenden Arbeit wurden die Items des analytischen Problemlösetests anders als die Items des KFT nicht zufällig auf verschiedene Skalen aufgeteilt, sondern den unterschiedlichen Arten von Problemstellungen zugeordnet. So lässt sich auch dieser Aspekt der faktoriellen Validität der Modellierung analytischer Problemlösekompetenz in den folgenden Analysen berücksichtigen. Alle Modelle wurden mit dem MLR-

Schätzalgorithmus und der Option „type=complex“ geschätzt, um der geclusterten Datenstruktur (Schülerinnen und Schüler geclustert in Schulen; siehe hierzu die Ausführungen weiter unten in diesem Abschnitt) und dem Problem fehlender Werte im KFT Rechnung zu tragen. Im Nested-Faktor-Modell (siehe Abbildung 3.1) werden die WL-Schätzer der drei Arten von Problemstellungen auf einen Faktor fluider Fähigkeiten  $ff'$  sowie auf einen um  $ff'$  residualisierten und damit von  $ff'$  unabhängigen Faktor der analytischen Problemlösekompetenz  $PI'$  zurückgeführt. Die beiden WL-Schätzer des KFT werden hingegen lediglich auf  $ff'$  zurückgeführt. Im erweiterten Standardmodell (siehe Abbildung 3.1) wurden die WL-Schätzer der drei Arten von Problemstellungen auf einen Faktor der analytischen Problemlösekompetenz  $PI$  zurückgeführt und die beiden WL-Schätzer des KFT auf einen mit  $PI$  korrelierten Faktor fluider Fähigkeiten  $ff$ . Dieses Modell wird als *erweitertes* Standardmodell bezeichnet, da wie im Standardmodell analytische Problemlösekompetenz und fluide Fähigkeiten als korrelierte Faktoren betrachtet werden, jedoch zusätzlich noch zwischen unterschiedlichen Arten von Problemstellungen differenziert wird. Das Einfaktormodell (siehe Abbildung 3.1) stellt das restriktivste Modell dar, bei dem alle WL-Schätzer auf einen Faktor allgemeiner kognitiver Fähigkeiten  $g$  zurückgeführt werden.

Da gezeigt werden kann, dass die beiden Faktoren fluider Fähigkeiten  $ff'$  im Nested-Faktor-Modell und  $ff$  im erweiterten Standardmodell mathematisch und empirisch äquivalent sind (z. B. Gustafsson, 1988; Sonnleitner et al., 2013; siehe auch Hartig & Höhler, 2008), erhalten beide die gleiche Bezeichnung und werden lediglich zum Zweck der besseren Differenzierbarkeit mit einem Apostroph unterschieden. So sind  $ff'$  und  $ff$  zwar konzeptionell deckungsgleich, unterscheiden sich aber in ihrer postulierten Wirkung. Während  $ff$  im erweiterten Standardmodell lediglich die Leistungen im KFT beeinflusst, hat  $ff'$  im Nested-Faktor-Modell sowohl Einfluss auf die Leistungen im KFT als auch auf die Leistungen im analytischen Problemlösetest. Im Gegensatz zu  $ff'$  und  $ff$  unterscheiden sich  $PI'$  und  $PI$  aber auch konzeptionell voneinander.  $PI$  repräsentiert eine analytische Problemlösekompetenz, die mit  $ff$  korreliert ist. Inwiefern sich die durch  $PI$  und  $ff$  repräsentierten latenten Fähigkeiten ähneln beziehungsweise unterscheiden, wird jedoch nicht näher spezifiziert. Demgegenüber repräsentiert  $PI'$  eine analytische Problemlösekompetenz, die unabhängig von  $ff'$  ist und ausschließlich bei den Leistungen im analytischen Problemlösetest zum Tragen kommt. Durch die Trennung von  $ff'$  und  $PI'$  im Nested-Faktor-Modell lässt sich also explizit

prüfen, inwiefern bei der Bearbeitung von analytischen Problemstellungen neben fluiden Fähigkeiten eine zusätzliche spezifische analytische Problemlösekompetenz notwendig ist. Im Nested-Faktor-Modell werden die Leistungen im Problemlösetest sowohl von  $fF'$  als auch von  $PI'$  beeinflusst, was der postulierten multiplen Bedingtheit von analytischen Problemlöseleistungen Rechnung trägt. Das Nested-Faktor-Modell sollte daher dem erweiterten Standardmodell überlegen sein und den besseren Modellfit aufweisen (Hypothese 1). Im Einfaktormodell wird im Gegensatz dazu weder der multiplen Bedingtheit von Problemlöseleistungen Rechnung getragen, noch werden analytische Problemlösekompetenz und fluide Fähigkeiten auf Konstruktebene konzeptionell getrennt. Sowohl das erweiterte Standardmodell als auch das Nested-Faktor-Modell sollten daher diesem theoretisch wenig plausiblen Modell überlegen sein und einen besseren Modellfit aufweisen.

Zur Bewertung des Modellfits wurden verschiedene Kriterien herangezogen. Zunächst wurde die Goodness-of-Fit- $\chi^2$ -Statistik bestimmt. Da bei dieser Statistik bei großen Stichproben selbst geringste Abweichungen der durch das Modell vorhergesagten Datenstruktur von der empirischen Datenstruktur signifikant werden (z. B. Byrne, 2012), sollten die entsprechenden Werte mit Vorsicht interpretiert werden. Des Weiteren wurden neben dem  $\chi^2$ -Differenzentest, der einen inferenzstatistischen Vergleich der Modellpassung verschiedener geschachtelter Modelle erlaubt (Satorra & Bentler, 2001), das AIC (Akaike Information Criterion) und das BIC (Bayesian Information Criterion) als informationstheoretische Maße der Modellpassung bestimmt (niedrigere Werte stehen jeweils für eine bessere Passung des Modells an die Daten; Kuha, 2004). Zudem wurden als weitere Fitstatistiken der CFI (Comparative Fit Index), der RMSEA (Root Mean Square Error of Approximation) sowie das SRMR (Standardized Root Mean Square Residual) berechnet. CFI Werte über .95, RMSEA-Werte unter .06 und SRMR-Werte unter .08 gelten dabei als Indikatoren für eine hinreichende Modellpassung (Hu & Bentler, 1999; vgl. Byrne, 2012).

Zur Beurteilung der inhaltlichen Interpretierbarkeit des im Nested-Faktor-Modell geschätzten Faktors der analytischen Problemlösekompetenz  $PI'$  (Hypothese 2) wurden, den Empfehlungen von Marsh und Grayson (1995) folgend, die Faktorvarianz sowie die Höhe der Faktorladungen der manifesten WL-Schätzer begutachtet. Demnach ist die inhaltliche Interpretierbarkeit gewährleistet, wenn eine signifikante Varianz und substantielle Faktorladungen (höher .30) geschätzt werden können (vgl. Carroll, 1993; Rindskopf & Rose, 1988).

Zur Bewertung der praktischen Bedeutsamkeit von  $PI'$  (Hypothese 3) wurden Zusammenhänge für  $PI'$ ,  $fF'$  (Nested-Faktor-Modell) sowie  $PI$  und  $fF$  (erweitertes Standardmodell) mit den Variablen Bildungsgang, Familiensprache, Geschlecht sowie den PISA-Testleistungen in Mathematik, Naturwissenschaften und Lesen (siehe PISA-Konsortium Deutschland, 2006) bestimmt. Die praktische Relevanz der Modellierung der analytischen Problemlösekompetenz mithilfe des Nested-Faktor-Modells ist dann gegeben, wenn sich für  $PI'$  zum einen signifikante Zusammenhänge mit den untersuchten externen Variablen zeigen und diese im Vergleich zu  $PI$  differenziell ausfallen (vgl. Hartig & Höhler, 2008; Hornung et al., 2011; Sonnleitner et al., 2013). Wäre dies nicht der Fall, würden  $PI'$  und  $PI$  zwar unterschiedliche Konzeptionen zugrunde liegen, diese hätten jedoch kaum praktische Relevanz. Die Analysen von Zusammenhängen mit externen Variablen wurden mit dem Programm HLM 7 (Raudenbush, Bryk, Cheong, Congdon & du Toit, 2011) durchgeführt, um der geclusterten Datenstruktur (Schülerinnen und Schüler geclustert in Schulen) gerecht zu werden. Hierzu wurden die Faktorwerte der spezifizierten Faktoren z-standardisiert und in HLM als Outcome-Variablen verwendet. Wie bereits in Studie 1 ausführlich dargestellt, ist davon auszugehen, dass die Ergebnisse von Schülerinnen und Schülern innerhalb eines Clusters (Schule) ähnlicher sind als die Ergebnisse im Vergleich zwischen den Clustern. Dies muss bei den Analysen berücksichtigt werden, um eine korrekte Signifikanzprüfung zu gewährleisten. Als Maß für die Ähnlichkeit der Schülerinnen und Schüler innerhalb einer Schule kann die Intraklassenkorrelation (ICC) bestimmt werden, die den Anteil der Gesamtvarianz der Leistungen angibt, der allein auf die Gruppenzugehörigkeit (Schule) der Schülerinnen und Schüler zurückgeführt werden kann. Wie sich zeigen lässt, können bereits geringe ICCs zu einer Inflation des alpha-Fehlers bei der Signifikanztestung führen (z. B. Goldstein, 2011). Analog zum Vorgehen in Studie 1 wurden für Mathematik, Naturwissenschaften und Lesen jeweils die fünf Plausible Values aus PISA 2003 verwendet, zu deren Schätzung neben den jeweiligen Testdaten auch zusätzliche Hintergrundinformationen über die Schülerinnen und Schüler herangezogen werden, und die dadurch eine hohe Messgenauigkeit besitzen (Carstensen, Knoll, Rost & Prenzel, 2004; OECD, 2005, Kap. 9). Diese Plausible Values wurden ebenfalls zunächst z-standardisiert. Anschließend wurde für jeden Plausible Value die Korrelation mit den Faktorwerten bestimmt, und diese fünf Werte wurden für jede Skala entsprechend gemittelt. Die Variablen Familiensprache und Geschlecht gingen als

dummy-kodierte Variablen in die Analysen ein. Die für diese Variablen berichteten Koeffizienten geben daher jeweils die Differenz der geschätzten Mittelwerte zwischen beiden Kategorien an. Die Bildungsgänge wurden ebenfalls dummy-kodiert. Hier geben die berichteten Koeffizienten jeweils die Mittelwertdifferenz zur Hauptschule an, die als Referenzkategorie gewählt wurde.

### 3.5 Ergebnisse

Tabelle 3.2 stellt die Ergebnisse zur Modellpassung der drei geschätzten Faktormodelle dar. Lediglich für das Nested-Faktor-Modell zeigte sich, indiziert durch einen nicht signifikanten Goodness-of-Fit  $\chi^2$ -Wert ( $p = .302$ ), eine gute Modellpassung an die Daten, während das erweiterte Standardmodell ( $p = .014$ ) und das Einfaktormodell ( $p < .001$ ) schlechte Modellpassungen aufwiesen. Auch wenn diese Werte aufgrund der großen Stichprobe nur mit Vorsicht zu interpretieren sind, stehen sie im Einklang mit den meisten anderen Fitstatistiken, die ebenfalls für eine bessere Modellpassung und damit für die Bevorzugung des Nested-Faktor-Modells gegenüber den anderen Modellen sprechen. So zeigte der  $\chi^2$ -Differenzentest für das gegenüber dem Nested-Faktor-Modell restriktivere erweiterte Standardmodell eine signifikante Verschlechterung der Modellpassung ( $\Delta\chi^2 = 10,10_{(2)}, p < .01$ ). Gleiches gilt für den Vergleich des Nested-Faktor-Modells mit dem noch restriktiveren Einfaktormodell ( $\Delta\chi^2 = 608,32_{(3)}, p < .001$ ). Auch der inferenzstatistische Vergleich des erweiterten Standardmodells mit dem Einfaktormodell (in Tabelle 3.2 nicht dargestellt) zeigte eine signifikant schlechtere Modellpassung des Einfaktormodells ( $\Delta\chi^2 = 4692,54_{(1)}, p < .001$ ). Der geringere Wert des AIC impliziert eine bessere Modellpassung des Nested-Faktor-Modells, während der BIC das erweiterte Standardmodell bevorzugt. Der CFI und das SRMR zeigen indessen für alle geschätzten Modelle gute Fitwerte, während der RMSEA nur für das Nested-Faktor-Modell und das erweiterte Standardmodell einen guten Modellfit anzeigt. Im Fall des RMSEA und des SRMR zeigen sich zudem die besten Fitwerte für das Nested-Faktor-Modell. Insgesamt sprechen die Koeffizienten der Modellpassung für eine bessere Passung des Nested-Faktor-Modelles gegenüber dem erweiterten Standardmodell und dem Einfaktormodell. Hypothese 1 konnte damit bestätigt werden.

**Tabelle 3.2: Anzahl geschätzter Parameter sowie Angaben zur Modellpassung des Nested-Faktor-Modells, des erweiterten Standardmodells sowie des Einfaktormodells**

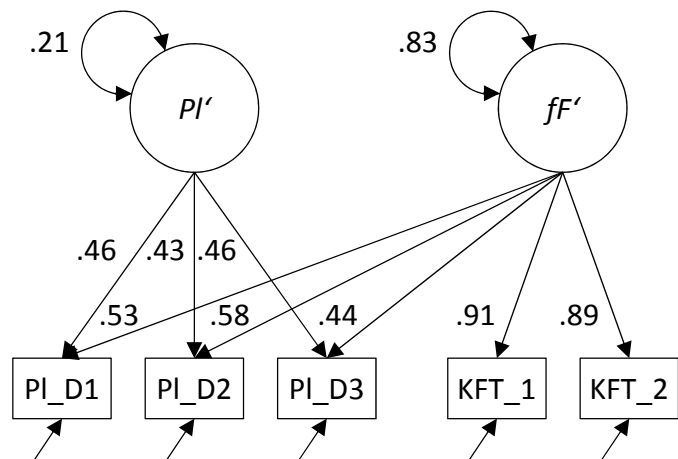
	Geschätzte Parameter	Goodness-of- Fit ( $p$ -Werte)	$\Delta\chi^2$ ( $df$ )	AIC	BIC	CFI	RMSEA	SRMR
Nested-Faktor-Modell	18	.302	-	27919	28023	1.000	0.009	0.002
Erw. Standardmodell	16	.014	10,10** (2)	27925	28017	1.000	0.030	0.009
Einfaktormodell	15	<.001	608,32*** (3)	28308	28394	0.992	0.210	0.061

*Anmerkungen.*  $\Delta\chi^2$ : Prüfgröße des  $\chi^2$ -Differenzentest (berechnet nach Satorra & Bentler, 2001) mit der Differenz der in beiden Modellen geschätzten Anzahl an Parametern als Freiheitsgrade ( $df$ ); AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; CFI: Comparative Fit Index, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual; \*\* $p < .01$ , \*\*\* $p < .001$ ;  $N = 2343$ .

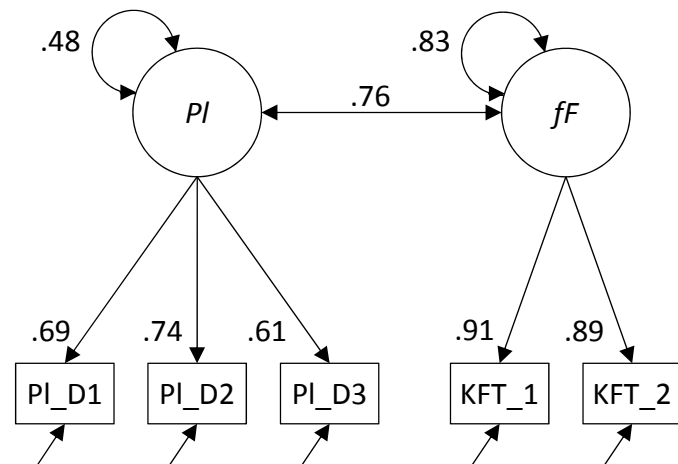
Abbildung 3.1 zeigt die standardisierten Koeffizienten der drei geschätzten Modelle. Beim Vergleich des Nested-Faktor-Modells mit dem erweiterten Standardmodell wird deutlich, dass beide Faktoren fluider Fähigkeiten  $fF'$  und  $fF$  identische Varianzen und Faktorladungen aufweisen. Dadurch wird empirisch nachgewiesen, dass beide Faktoren konzeptionell deckungsgleich sind. Der Vergleich der Faktorladungen und Varianzen der beiden Faktoren der analytischen Problemlösekompetenz  $PI'$  und  $PI$  macht demgegenüber deutlich, dass es sich bei diesen Faktoren um konzeptionell unterschiedliche Faktoren handelt. Für den um  $fF'$  residualisierten Faktor  $PI'$  im Nested-Faktor-Modell zeigten sich niedrigere Faktorladungen als für  $PI$  im erweiterten Standardmodell, allerdings lagen die Faktorladungen auch im Nested-Faktor-Modell über .30. Ebenso wurde für  $PI'$  mit  $\sigma^2 = .21$  ( $p < .001$ ) eine niedrigere Varianz geschätzt als für  $PI$  ( $\sigma^2 = .48$ ;  $p < .001$ ), die aber dennoch signifikant von 0 verschieden war. Der um  $fF'$  residualisierte Faktor der analytischen Problemlösekompetenz  $PI'$  im Nested-Faktor-Modell besitzt somit empirische Substanz und kann inhaltlich interpretiert werden. Hypothese 2 konnte damit bestätigt werden.



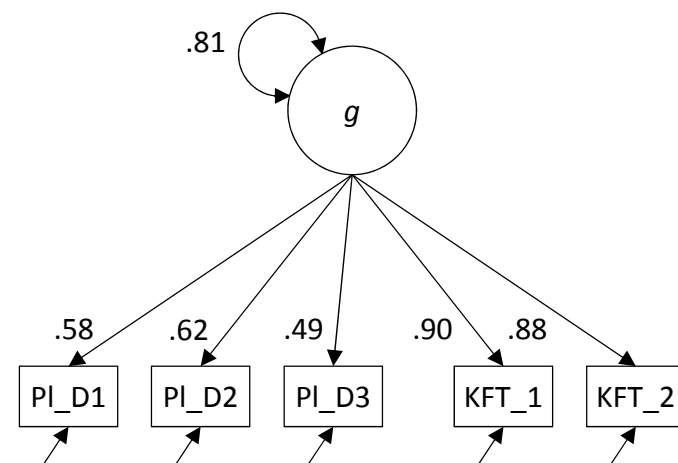
### Nested-Faktor-Modell



### Erweitertes Standardmodell



### Einfaktormodell



**Abbildung 3.1: Varianzen, Faktorladungen und Korrelationen für das Nested-Faktor-Modell, das erweiterte Standardmodell und das Einfaktormodell (standardisierte Koeffizienten; für alle Koeffizienten gilt:  $p < .001$ )**

In Tabelle 3.3 sind die Ergebnisse der Untersuchung von Zusammenhängen der Faktorwerte des Nested-Faktor-Modells und des erweiterten Standardmodells mit externen Variablen dargestellt. Die nahezu identischen Koeffizienten für  $fF$  und  $fF'$  machen auch hier deutlich, dass beide Faktoren konzeptionell deckungsgleich sind. Die Notwendigkeit einer mehrebenenanalytischen Betrachtung wird ersichtlich an den für jeden Faktor substantziellen Anteilen an der jeweiligen Gesamtvarianz der Personenparameter auf Schulebene (erfasst über die ICC). Für den Faktor  $PI'$  zeigte sich eine substantzielle Intraklassenkorrelation von .16. Ebenso ergaben sich für  $PI'$  signifikante Mittelwertunterschiede zwischen den verschiedenen Bildungsgängen (jeweils in Relation zur Hauptschule) mit Ausnahme des Vergleichs zwischen Gesamtschule und Realschule, da sich hier die entsprechenden Konfidenzintervalle überschneiden. Insgesamt erklärte die Variable Bildungsgang 66 % der Varianz auf Schulebene des Faktors  $PI'$ . Im Hinblick auf die Familiensprache und das Geschlecht zeigten sich signifikante Mittelwertunterschiede zugunsten der Schülerinnen und Schüler mit Deutsch als Familiensprache (0.451,  $KI_{95} = 0.260/0.643$ ) sowie zuungunsten der Mädchen (-.118,  $KI_{95} = -0.194/-0.042$ ). Darüber hinaus zeigten sich signifikante Korrelationen mit den fachbezogenen Kompetenzen in der Mathematik, den Naturwissenschaften und im Lesen. So ließen sich für den um  $fF'$  residualisierten Faktor  $PI'$  neben substantziellen Faktorladungen und einer signifikanten Varianz ebenfalls signifikante Zusammenhänge mit den untersuchten externen Variablen nachweisen.

Im Vergleich der Faktoren  $PI'$  und  $PI$  zeigten sich Hinweise auf differenzielle Zusammenhänge mit den untersuchten Variablen. So lagen für  $PI'$  ein geringerer Anteil der Gesamtvarianz auf Schulebene (16 % versus 44 %), die Mittelwertunterschiede zwischen Realschule und Hauptschule sowie zwischen Gymnasium und Hauptschule fielen geringer aus, und der Bildungsgang erklärte weniger Varianz auf Schulebene. Zwar fiel auch der Mittelwertunterschied zugunsten der Schülerinnen und Schüler mit Deutsch als Familiensprache deskriptiv geringer sowie der Mittelwertunterschied zuungunsten der Mädchen für  $PI'$  deskriptiv größer aus als für  $PI$ , allerdings überschneiden sich die jeweiligen Konfidenzintervalle, sodass der Unterschied nicht inferenzstatistisch abgesichert werden konnte.<sup>6</sup> Signifikante Unterschiede zeigten sich jedoch hinsichtlich der Korrelationen mit fachbezogenen

---

<sup>6</sup> Es sei darauf hingewiesen, dass die dargestellten Ergebnisse keine generalisierbaren Aussagen auf die Population aller 15-Jährigen Schülerinnen und Schüler in Deutschland zulassen. Da die in PISA gezogene

Kompetenzen. Hier fielen die Korrelationen für  $PI'$  durchweg geringer aus als für  $PI$ . Aufgrund der signifikanten Zusammenhänge von  $PI'$  mit den untersuchten externen Variablen, die sich, wenn auch nicht für alle untersuchten Variablen, signifikant von denen für  $PI$  unterschieden, konnte Hypothese 3, wenn auch nur mit Einschränkungen, bestätigt werden.

---

Gesamtstichprobe keine repräsentative Zufallsstichprobe darstellt, ist für Aussagen auf Populationsebene eine entsprechende Gewichtung notwendig, die für die hier verwendete Teilstichprobe nicht durchgeführt wurde. Dies erklärt die Unterschiede der hier dargestellten Ergebnisse zu Geschlechtsunterschieden im analytischen Problemlösen, von den bei Leutner et al (2004) berichteten.

**Tabelle 3.3: Intraklassenkorrelationen (ICC) und Mittelwertdifferenzen der Faktorwerte des erweiterten Standardmodells und des Nested-Faktor-Modells nach Bildungsgang, Familiensprache und Geschlecht sowie Korrelationen mit fachbezogenen Kompetenzen (95%-Konfidenzintervalle in eckigen Klammern)**

	Erweitertes Standardmodell		Nested-Faktor-Modell	
	<i>PI</i>	<i>fF</i>	<i>PI'</i>	<i>fF'</i>
ICC	.44	.35	.16	.35
Bildungsgang				
Gesamtschule	0.474[0.248/0.700]***	0.410[0.195/0.626]***	0.295[0.090/0.500]**	0.408[0.194/0.623]***
Realschule	0.782[0.626/0.938]***	0.690[0.547/0.834]***	0.448[0.304/0.592]***	0.690[0.546/0.833]***
Gymnasium	1.549[1.416/1.681]***	1.372[1.244/1.500]***	0.862[0.735/0.988]***	1.371[1.244/1.499]***
<i>R</i> <sup>2</sup>	.79	.80	.66	.80
Familiensprache (1 = Deutsch)	0.554[0.411/0.697]***	0.498[0.317/0.678]***	0.451[0.260/0.643]***	0.495[0.314/0.675]***
Geschlecht (1 = weiblich)	-0.065[-0.127/-0.003]*	0.002[-0.066/0.071]	-0.118[-0.194/-0.042]***	0.004[-0.065/0.072]
Mathematik	.802[.778/.826]***	.707[.676/.739]***	.448[.407/.489]***	.707[.676/.738]***
Naturwissenschaften	.731[.699/.763]***	.640[.606/.675]***	.439[.397/.481]***	.640[.606/.674]***
Lesen	.711[.676/.745]***	.622[.583/.661]***	.428[.387/.468]***	.622[.583/.661]***

*Anmerkungen.* Geschätzte Mittelwertdifferenzen der z-standardisierten Faktorwerte für Bildungsgang (Referenzkategorie: Hauptschule), Familiensprache und Geschlecht sowie Korrelationen mit Mathematik, Naturwissenschaften und Lesen (für die Domänen Mathematik, Naturwissenschaften und Lesen wurden die Faktorwerte mit jedem der fünf Plausible Values pro Domäne aus PISA 2003 [OECD, 2005] korreliert und diese Korrelationen im Anschluss für jede Domäne gemittelt); \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ ;  $N = 2343$ .

## 3.6 Diskussion

### 3.6.1 Zusammenfassung der Ergebnisse

Ziel von Studie 2 war die Untersuchung von Aspekten der diskriminanten Validität der Modellierung analytischer Problemlösekompetenz gegenüber fluiden Fähigkeiten als zentralen kognitiven Sekundärfähigkeiten der allgemeinen kognitiven Fähigkeiten (Carroll, 1993). Hierzu wurden ein Einfaktormodell, ein erweitertes Standardmodell und ein Nested-Faktor-Modell miteinander verglichen sowie Zusammenhänge der Faktorwerte des erweiterten Standardmodells sowie des Nested-Faktor-Modells mit externen Variablen bestimmt. Hypothese 1, nach der das Nested-Faktor-Modell eine bessere Modellpassung aufweisen sollte als das Standardmodell, und Hypothese 2, der zufolge der Faktor analytischer Problemlösekompetenz im Nested-Faktor-Modell substantiell und inhaltlich interpretierbar sein sollte, also eine signifikante Varianz und substantielle Faktorladungen aufweisen sollte, konnten bestätigt werden. Hypothese 3, nach welcher der Faktor analytischer Problemlösekompetenz im Nested-Faktor-Modell praktische Bedeutsamkeit aufweisen sollte, sich also differenzielle Zusammenhänge dieses Faktors im Vergleich zum entsprechenden Faktor im erweiterten Standardmodell mit externen Variablen zeigen sollten, konnte teilweise bestätigt werden.

### 3.6.2 Theoretische und praktische Implikationen

In bisherigen Untersuchungen auf Basis des Standardmodells wurden Gemeinsamkeiten und Unterschiede zwischen analytischer Problemlösekompetenz und fluiden Fähigkeiten nicht explizit spezifiziert (z. B. Leutner, Wirth et al., 2005). Das in der vorliegenden Studie 2 favorisierte Nested-Faktor-Modell leistet genau dies, indem die Leistungsvarianz der analytischen Problemlöseskalen in zwei voneinander unabhängige Faktoren zerlegt wird. Die Ergebnisse der durchgeführten Modellvergleiche unterstützen die theoretische Annahme einer von fluiden Fähigkeiten unterscheidbaren analytischen Problemlösekompetenz (z. B. Leutner, Funke et al., 2005). Die Ergebnisse stehen damit im Einklang mit bisherigen Untersuchungen zur diskriminanten Validität der Modellierung analytischer Problemlösekompetenz gegenüber fluiden Fähigkeiten (z. B. Fischer et al., 2015; Leutner et al.

2004; Scherer & Tiemann, 2014; Wirth et al., 2005). Sie erweitern diese aber dahingehend, dass gezeigt werden kann, dass analytische Problemlösekompetenz und fluide Fähigkeiten nicht nur auf Konstruktebene als hoch korreliert konzeptualisiert werden müssen, sondern sich auch eine von fluiden Fähigkeiten unabhängige analytische Problemlösekompetenz empirisch etablieren lässt. So verbleibt im Nested-Faktor-Modell selbst unter Kontrolle der fluiden Fähigkeiten Leistungsvarianz der analytischen Problemlöseskalen, die auf einen von fluiden Fähigkeiten unabhängigen, substantiellen und inhaltlich interpretierbaren analytischen Problemlösefaktor zurückgeführt werden kann.

Das aufgrund der vorliegenden Ergebnisse favorisierte Nested-Faktor-Modell stellt darüber hinaus eine theoretisch adäquatere Modellierung der analytischen Problemlösekompetenz dar, indem es im Gegensatz zum Standardmodell die Annahme der multiplen Bedingtheit von analytischen Problemlöseleistungen explizit aufnimmt und dadurch eine empirische Prüfung derselben erlaubt. Die vorliegenden Ergebnisse auf Basis des Nested-Faktor-Modells stützen diese Annahme, indem gezeigt werden konnte, dass die analytischen Problemlöseaufgaben sowohl substantielle Ladungen auf dem Faktor fluider Fähigkeiten als auch auf dem Problemlösefaktor aufweisen. Das Lösungsverhalten in Aufgaben der Art *Entscheidungen treffen* ( $\alpha_{(P1\_D1)} = .53$ ) und in Aufgaben der Art *Systeme analysieren und entwerfen* ( $\alpha_{(P1\_D2)} = .58$ ) scheint dabei in stärkerem Maße durch fluide Fähigkeiten bedingt zu sein als Aufgaben der Art *Fehler suchen* ( $\alpha_{(P1\_D3)} = .44$ ; siehe Abbildung 3.1). Dieser Befund deckt sich auch mit den Ergebnissen aus Studie 1. Wie Tabelle 3.3 zu entnehmen ist, zeigen sich für Aufgaben der Art *Systeme analysieren und entwerfen* darüber hinaus auch stärkere Zusammenhänge mit fachbezogenen Kompetenzen im Vergleich zu den beiden anderen Arten von Problemstellungen, was den Schluss nahe legt, dass dies vor allem auf die mit fluiden Fähigkeiten gemeinsame Varianz dieser Art von Problemstellungen zurückgeführt werden kann. Fluide Fähigkeiten alleine reichen jedoch nicht aus, um die Leistungsvarianz der analytischen Problemlöseaufgaben zu erklären. Das Lösungsverhalten in diesen Aufgaben wird ebenfalls durch einen von fluiden Fähigkeiten unabhängigen Faktor analytischer Problemlösekompetenz bedingt. Dies trifft auf die drei Arten von Problemstellungen in etwa gleichem Ausmaß zu, wie die nahezu identischen Ladungen auf dem analytischen Problemlösefaktor zeigen (siehe Abbildung 3.1). Die Ergebnisse von Studie 2 wider-

legen damit einmal mehr die Behauptungen von Rindermann (2006), wonach unter anderem für die Modellierung analytischer Problemlösekompetenz in PISA 2003 keine diskriminante Validität gegenüber Maßen der Intelligenz, insbesondere dem schlussfolgernden Denken und den fluiden Fähigkeiten, angenommen werden könne (für Befunde zur diskriminanten Validität von Maßen dynamischer bzw. komplexer Problemlösekompetenz gegenüber Maßen der Intelligenz siehe z. B. Greiff, Wüstenberg, Molnár, Fischer, Funke & Csapó, 2013; Sonnleitner et al., 2012; Wüstenberg, Greiff & Funke, 2012).

Hinweise zur praktischen Bedeutsamkeit der unterschiedlichen Modellierungen analytischer Problemlösekompetenz auf Basis des erweiterten Standardmodells und des Nested-Faktor-Modells sind Tabelle 3.3 zu entnehmen. Hier zeigt sich, dass je nach Art der Modellierung analytischer Problemlösekompetenz unterschiedliche Befunde hinsichtlich des Zusammenhangs mit externen Variablen resultieren können. Dies verdeutlicht die unterschiedlichen Konzeptionen der analytischen Problemlösekompetenz in beiden Modellen, einmal als einer mit fluiden Fähigkeiten konfundierten Kompetenz im erweiterten Standardmodell und einmal als einer von fluiden Fähigkeiten unabhängigen Kompetenz im Nested-Faktor-Modell. So ist die mit fluiden Fähigkeiten konfundierte analytische Problemlösekompetenz im erweiterten Standardmodell stärker geclustert als die um fluide Fähigkeiten residualisierte analytische Problemlösekompetenz im Nested-Faktor-Modell ( $ICC_{PI'} = .16$  versus  $ICC_{PL} = .44$ ). Es liegt also ein größerer Anteil der jeweiligen Gesamtvarianz der analytischen Problemlösekompetenz auf Schulebene. Dies bedeutet, dass sich Schülerinnen und Schüler einer Schule hinsichtlich der mit fluiden Fähigkeiten konfundierten analytischen Problemlösekompetenz stärker ähneln als bei der um fluide Fähigkeiten bereinigten analytischen Problemlösekompetenz. Zudem wird bei der, um fluide Fähigkeiten bereinigten, analytischen Problemlösekompetenz ein geringerer Anteil der Varianz auf Schulebene durch den Bildungsgang erklärt, als dies bei der mit fluiden Fähigkeiten konfundierten analytischen Problemlösekompetenz der Fall ist. Interessant ist darüber hinaus, dass der Anteil der Varianz auf Schulebene bei der analytischen Problemlösekompetenz im Nested-Faktor-Modell deutlich unter dem der fluiden Fähigkeiten liegt, während er im erweiterten Standardmodell etwas darüber liegt. Es scheint also vor allem der auf fluide Fähigkeiten zurückzuführende Anteil an der Leistungsvarianz in analytischen Problemlöseauf-

gaben zu sein, der für die geclusterte Datenstruktur, also für die Abhängigkeit der analytischen Problemlösekompetenz von der Schulzugehörigkeit der Schülerinnen und Schüler, verantwortlich ist.

Besonders deutlich werden die unterschiedlichen Implikationen der beiden Modellierungen analytischer Problemlösekompetenz, wenn die jeweiligen Zusammenhänge mit fachbezogenen Kompetenzen betrachtet werden. Bei der mit fluiden Fähigkeiten konfundierten analytischen Problemlösekompetenz liegen die Zusammenhänge mit fachbezogenen Kompetenzen in der etwa gleichen Größenordnung wie die entsprechenden Zusammenhänge für fluide Fähigkeiten. Demgegenüber sind die entsprechenden Zusammenhänge für die um fluide Fähigkeiten residualisierte analytische Problemlösekompetenz signifikant niedriger. Es scheint also, in einem nicht unerheblichen Maße, der auf fluide Fähigkeiten zurückzuführende Anteil an der Leistungsvarianz in analytischen Problemlöseaufgaben zu sein, der für die vergleichsweise hohen Korrelationen zwischen analytischer Problemlösekompetenz und fachbezogenen Kompetenzen verantwortlich ist. Allerdings – und dies liefert Argumente für einen weiteren Aspekt der Konstruktvalidität – zeigen sich auch für die um fluide Fähigkeiten residualisierte (also davon unabhängige) analytische Problemlösekompetenz substanzielle Zusammenhänge mit fachbezogenen Kompetenzen. Hieran wird der Mehrwert der analytischen Problemlösekompetenz bei der Vorhersage fachbezogener Kompetenzen, (*inkrementelle Validität*; Haynes & Lench, 2003) gegenüber fluiden Fähigkeiten deutlich. Dieser Aspekt der Konstruktvalidität der Modellierung einer Kompetenz wird auch als angewandte Form der Validität bezeichnet, da sie insbesondere in der individualdiagnostischen Praxis eine Rolle spielt, wenn es um die Frage geht, welche zusätzlichen Instrumente eingesetzt werden sollen, um die Qualität einer Vorhersage zur Vorbereitung von Entscheidungen auf Einzelfallebene zu verbessern (Hunsley & Meyer, 2003). Die inkrementelle Validität der Modellierung analytischer Problemlösekompetenz, die an dieser Stelle lediglich am Rande und in einem querschnittlichen Design betrachtet wurde, soll in der folgenden Studie 3 in einem längsschnittlichen Design näher untersucht werden.

Zusammenfassend lässt sich festhalten, dass die Modellierung der analytischen Problemlösekompetenz mithilfe des Nested-Faktor-Modells eine bessere Anpassung an die



Daten zeigt und damit geeignet ist, die Zusammenhänge zwischen analytischer Problemlösekompetenz und fluiden Fähigkeiten adäquater zu beschreiben als bisherige Modellierungen auf Basis des Standardmodells. Das Nested-Faktor-Modell liefert weitere Belege für die diskriminante und die inkrementelle Validität der Modellierung analytischer Problemlösekompetenz gegenüber fluiden Fähigkeiten sowie für die multiple Bedingtheit von analytischen Problemlöseleistungen. Es liefert damit wichtige Informationen für ein differenzierteres Bild der analytischen Problemlösekompetenz und der Rolle von fluiden Fähigkeiten bei der Erklärung von Leistungen von Schülerinnen und Schülern bei der Bearbeitung analytischer Problemstellungen.

### 3.6.3 Grenzen und Ausblick

Zur Untersuchung von Zusammenhängen der analytischen Problemlösekompetenz im erweiterten Standardmodell und im Nested-Faktor-Modell mit externen Variablen wurde, analog zu Studie 1, ein sequenzielles Vorgehen gewählt. Dabei wurden zunächst Messmodelle spezifiziert und anschließend die Zusammenhänge der Faktorwerte mit externen Variablen sowie Mittelwertdifferenzen für verschiedene Personengruppen bestimmt. Dabei wurde der geclusterten Struktur der verwendeten Daten Rechnung getragen, allerdings ermöglicht dieses Vorgehen keine messfehlerfreie Modellierung auf latenter Ebene (für ein ähnliches Vorgehen siehe z. B. Hartig & Höhler, 2008; vgl. Sonnleitner et al., 2013), wie dies durch entsprechende Strukturgleichungsmodelle möglich wäre. Wie bereits in Studie 1, wurde auch in Studie 2 auf die dafür notwendige, umfangreiche Spezifizierung von Messmodellen für die Tests zur Erfassung der fachbezogenen Kompetenzen und damit die komplette Neuskalierung der verwendeten PISA 2003 Daten verzichtet. Vor dem Hintergrund, dass mit den Plausible Values aus PISA 2003 bereits methodisch elaborierte Maße der fachbezogenen Kompetenzskalen vorliegen, scheint dieses Vorgehen gerechtfertigt.

Eine weitere Einschränkung der Studie 2 ergibt sich aus der Schwierigkeit die um fluide Fähigkeiten residualisierte analytische Problemlösekompetenz im Nested-Faktor-Modell inhaltlich zu beschreiben. Die Faktorladungen der drei Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen*

liefern hier kaum inhaltliche Anhaltspunkte, da sie ein nahezu identisches Ladungsmuster aufweisen. Somit kann über die inhaltliche Bedeutung dieses Faktors an dieser Stelle nur spekuliert werden. Eine Rolle könnte dabei die Fähigkeit spielen, Wissen aus unterschiedlichen Domänen heranzuziehen sowie die Fähigkeit, bei der Bearbeitung komplexer Problemstellungen systematisch, strategisch und selbstreguliert vorzugehen (Leutner, Fleischer & Wirth, 2006). Diese Anforderungen, die insbesondere auf konditionales Wissen abzielen, werden bei den Testitems der Subskala *N2 Figurenanalogien* des KFT, die weniger komplex und unter einem gewissen Zeitdruck zu bearbeiten sind, nicht gestellt. Allerdings lassen die in Studie 2 verwendeten Daten hierzu nur Spekulationen zu.

Solange es keine Items gibt, die dazu geeignet sind, die von fluiden Fähigkeiten unabhängige analytische Problemlösekompetenz, wie sie im Nested-Faktor-Modell spezifiziert wird, direkt zu operationalisieren, also nur auf diesem Faktor laden, bleibt die praktische Bedeutsamkeit dieser Modellierung für den Bereich der Kompetenzdiagnostik als Spezialfall der Individualdiagnostik zunächst eingeschränkt. Für zukünftige Forschungsarbeiten ist es daher zunächst notwendig, theoriegeleitet weitere Items zur Erfassung analytischer Problemlösekompetenz zu konstruieren und diese systematisch in ihren konzeptuellen Merkmalen hinsichtlich Nähe und Distanz zu fluiden Fähigkeiten zu variieren. In einem ersten Schritt kämen hierfür fächerunabhängige Problemstellungen in Frage, wie sie in PISA 2000 eingesetzt wurden (Klieme, Hartig & Wirth, 2005) und von denen angenommen werden kann, dass sie eine größere Nähe zu fluiden Fähigkeiten aufweisen als die im Rahmen von PISA 2003 eingesetzten fächerverbindenden Problemstellungen.

### 3.7 Literatur

- Adams, R. J., Wilson, M. R. & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1–23. doi: 10.1177/0146621697211001
- Baumert, J., Brunner, M., Lüdtke, O. & Trautwein, U. (2007). Was messen internationale Schulleistungsstudien? – Resultate kumulativer Wissenserwerbsprozesse. Eine Antwort auf Heiner Rindermann. *Psychologische Rundschau*, 58, 118–128. doi: 10.1026/0033-3042.58.2.118
- Baumert, J., Lüdtke, O., Trautwein, U. & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review*, 4, 165–176. doi: 10.1016/j.edurev.2009.04.002
- Bollen, K. A. & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305–314. doi: 10.1037/0033-2909.110.2.305
- Brunner, M. (2008). No g in education? *Learning and Individual Differences*, 18, 152–165. doi: 10.1016/j.lindif.2007.08.005
- Brunner, M., Krauss, S. & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence*, 36, 403–421. doi: 10.1016/j.intell.2007.11.002
- Buchwald, F., Fleischer, J., Rumann, S., Wirth, J. & Leutner, D. (2017). Training in components of problem-solving competence: An experimental study of aspects of the cognitive potential exploitation hypothesis. In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 315–331). Berlin: Springer. doi: 10.1007/978-3-319-50030-0\_19
- Byrne, B. M. (2012). *Structural equation modeling with Mplus. Basic concepts, applications, and programming*. New York, NY: Routledge.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. doi: 10.1037/h0046016

- Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research*, 36, 589–609. doi: 10.1207/S15327906MBR3604\_05
- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Carstensen, C. H., Knoll, S., Rost, J. & Prenzel, M. (2004). Technische Grundlagen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 371–387). Münster: Waxmann.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22. doi: 10.1037/h0046743
- Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, 63, 453–482. doi: 10.1146/annurev-psych-120710-100353
- Eysenck, H. J. (2004). *Die IQ-Bibel. Intelligenz verstehen und messen*. Stuttgart: Klett.
- Fischer, A., Greiff, S., Wüstenberg, S., Fleischer, J., Buchwald, F. & Funke, J. (2015). Assessing analytic and interactive aspects of problem solving competency. *Learning and Individual Differences*, 39, 172–179. doi: 10.1016/j.lindif.2015.02.008
- Fleischer, J., Koeppen, K., Kenk, M., Klieme, E. & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. *Zeitschrift für Erziehungswissenschaft*, 16 (suppl. 1), 5–22.
- Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester, UK: Wiley.
- Gottfredson, L. S. (2003). g, jobs and life. In H. Nyborg (Ed.), *The scientific study of general intelligence. Tribute to Arthur R. Jensen* (pp. 293–342). Oxford, UK: Pergamon.
- Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 35–72). Hillsdale, NJ: Erlbaum.
- Gustafsson, J.-E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407–434. doi: 10.1207/s15327906mbr2804\_2

- Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J. & Csapó, B. (2013). Complex problem solving in educational contexts—Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, 105, 364–379. doi: 10.1037/a0031856
- Hartig, J. & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. *Zeitschrift für Psychologie / Journal of Psychology*, 216, 89–101. doi: 10.1027/0044-3409.216.2.89
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer. doi: 10.1007/3-540-33020-8\_9
- Haynes, S. N. & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, 15, 456–466. doi: 10.1037/1040-3590.15.4.456
- Heller, K. A. & Perleth, C. (2000). *KFT 4-12+R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Hogrefe.
- Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443–451). New York, NY: Macmillan.
- Horn, J. L. & Masunaga, H. (2006). A merging theory of expertise and intelligence. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *Cambridge handbook on expertise and expert performance* (pp. 587–612). Cambridge, UK: University Press. doi: 10.1017/CBO9780511816796.034
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan & J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.
- Hornung, C., Brunner, M., Reuter, R. A. P. & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence*, 39, 210–221. doi: 10.1016/j.intell.2011.03.002
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. doi: 10.1080/10705519909540118

- Hunsley, J. & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological and statistical issues. *Psychological Assessment*, 15(4), 446–455. doi: 10.1037/1040-3590.15.4.446
- Klieme, E. (2004). Assessment of cross-curricular problem-solving competencies. In J. H. Moskowitz & M. Stephens (Eds.), *Comparing learning outcomes. International assessments and education policy* (pp. 81–107). London: Routledge.
- Klieme, E., Funke, J., Leutner, D., Reimann, P. & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47, 179–200.
- Klieme, E., Hartig, J. & Wirth, J. (2005). Analytisches Problemlösen: Messansatz und Befunde zu Planungs- und Entscheidungsaufgaben. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 37–54). Wiesbaden: VS.
- Koeppen, K., Hartig, J., Klieme, E. & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie / Journal of Psychology*, 216, 61–73. doi: 10.1027/0044-3409.216.2.61
- Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods and Research*, 33, 188–229. doi: 10.1177/0049124103262065
- Kyllonen, P. C. & Lee, S. (2005). Assessing problem solving in context. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 11–25). Thousand Oaks, CA: Sage. doi: 10.4135/9781452233529
- Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. *Computers in Human Behavior*, 18, 685–697. doi: 10.1016/S0747-5632(02)00024-9
- Leutner, D., Fleischer, J. & Wirth, K. (2006). Problemlösekompetenz als Prädiktor für zukünftige Kompetenz in Mathematik und in den Naturwissenschaften. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 119–137). Münster: Waxmann.

- Leutner, D., Funke, J., Klieme, E. & Wirth, J. (2005). Problemlösefähigkeit als fächerübergreifende Kompetenz. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 11–19). Wiesbaden: VS.
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147–175). Münster: Waxmann.
- Leutner, D., Wirth, J., Klieme, E. & Funke, J. (2005). Ansätze zur Operationalisierung und deren Erprobung im Feldtest zu PISA 2000. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 21–36). Wiesbaden: VS.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R* (2. Aufl.). Göttingen: Hogrefe.
- Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151–173. doi: 10.1207/S15328007SEM0902\_1
- Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford, UK: University Press.
- Marsh, H.W. & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 177–198). Thousand Oaks, CA: Sage.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1–10. doi: 10.1016/j.intell.2008.08.004
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. doi: 10.1037/0003-066X.50.9.741
- Muthén, L. K. & Muthén, B. O. (1998–2012). *Mplus user's guide (7th edition)*. Los Angeles, CA: Muthén & Muthén.

- OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing.
- OECD. (2004). *Problem solving for tomorrow's world: First measurements of cross-curricular competencies from PISA 2003*. Paris: OECD Publishing.
- OECD. (2005). *PISA 2003: Technical report*. Paris: OECD Publishing.
- OECD. (2012). *PISA 2009: Technical report*. Paris: OECD Publishing.
- PISA-Konsortium Deutschland (Hrsg.). (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Prenzel, M., Drechsel, B., Carstensen, C. H. & Ramm, G. (2004). PISA 2003 – eine Einführung. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 13–46). Münster: Waxmann.
- Prenzel, M., Walter, O. & Frey, A. (2007). PISA misst Kompetenzen. Eine Replik auf Rindermann (2006). Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, 58, 128–136. doi: 10.1026/0033-3042.58.2.128
- Raaheim, K. (1988). Intelligence and task novelty. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 73–97). Hillsdale, NJ: Erlbaum.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T. & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Chicago, IL: SSI.
- Ree, M. J. & Carretta, T. R. (2002). g2K. *Human Performance*, 15, 3–23. doi: 10.1080/08959285.2002.9668081
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? *Psychologische Rundschau*, 57, 69–86.
- Rindskopf, D. & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, 23, 51–67. doi: 10.1207/s15327906mbr2301\_3
- Rost, D. H. (2009). *Intelligenz Fakten und Mythen*. Weinheim: Beltz.



- Satorra, A. & Bentler, P.M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514. doi: 10.1007/BF02296192
- Scherer, R. & Tiemann, R. (2014). Evidence on the effects of task interactivity and grade level on thinking skills involved in complex problem solving. *Thinking Skills and Creativity*, 11, 48–64. doi: 10.1016/j.tsc.2013.10.003
- Schulze, R. (2005). Modeling structures of intelligence. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 241–263). Thousand Oaks, CA: Sage.
- Snyderman, M. & Rothman, C. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, 42, 137–144. doi: 10.1037/0003-066X.42.2.137
- Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R. et al. (2012). The Genetics Lab: Acceptance and psychometric characteristics of a computer-based microworld assessing complex problem solving. *Psychological Test and Assessment Modeling*, 54, 54–72.
- Sonnleitner, P., Keller, U., Martin, R. & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41, 289–305. doi: 10.1016/j.intell.2013.05.002
- Spearman, C. (1927). *The abilities of man. Their nature and measurement*. New York, NY: Macmillan.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.
- Sternberg, R. J. (2012). The triarchic theory of successful intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (3rd ed., pp. 103–177). New York, NY: Guilford Press.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450. doi: 10.1007/BF02294627

- Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373–392). Thousand Oaks, CA: Sage.
- Wilhelm, O. & Engle, R. W. (Eds.). (2005). *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage.
- Wirth, J., Leutner, D. & Klieme, E. (2005). Problemlösekompetenz – Ökonomisch und zugleich differenziert erfassbar? In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 73–82). Wiesbaden: VS.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell, Australia: ACER Press.
- Wüstenberg, S., Greiff, S. & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence*, 40, 1–14. doi: 10.1016/j.intell.2011.11.003

## 4 STUDIE 3: PROGNOSTISCHE UND INKREMENTELLE VALIDITÄT DER MODELLIERUNG ANALYTISCHER PROBLEMLÖSEKOMPETENZ – VORHERSAGE ZUKÜNFTIGER FACHBEZOGENER KOMPETENZEN

### 4.1 Einleitung

In den beiden vorherigen Kapiteln 2 und 3 standen Fragen nach der Binnenstruktur der analytischen Problemlösekompetenz (*faktorielle Validität*) und nach der Abgrenzbarkeit von analytischer Problemlösekompetenz und fluiden Fähigkeiten (*diskriminante Validität*) jeweils in querschnittlichen Designs im Fokus. In Studie 3 werden nun mit der prognostischen und der inkrementellen Validität der Modellierung analytischer Problemlösekompetenz zwei weitere Aspekte der Validität in den Blick genommen.

Der Aufbau von Problemlösekompetenz ist in zweierlei Hinsicht bedeutsam für Bildungssysteme: zum einen als direktes Ziel von Schule und Unterricht (z. B. Binkley et al., 2012; Blum, Drücke-Noe, Hartung & Köller, 2006; OECD, 2004, 2014) und zum anderen, weil zu vermuten ist, dass Problemlösekompetenz eine wichtige Grundlage für weiteres schulisches Lernen darstellt (Leutner, Klieme, Meyer & Wirth, 2004; OECD, 2004, 2014). Der zuletzt genannte Aspekt betrifft die Frage, wie *fächerübergreifende* analytische Problemlösekompetenz mit unterschiedlichen *fachbezogenen* Kompetenzen zusammenhängt und welchen Beitrag sie zur Vorhersage dieser leisten kann (*prognostische Validität*; Cronbach & Gleser, 1965; siehe auch Hartig, Frey & Jude, 2012; *externer Aspekt* der Konstruktvalidität, sensu Messick, 1995). Vor dem Hintergrund der in Abschnitt 1.3.3 dargestellten Potenzialausschöpfungshypothese, der zufolge Schülerinnen und Schüler in Deutschland über kognitives Potenzial verfügen, welches sich beim analytischen Problemlösen manifestiert, in der Schule jedoch nicht hinreichend zum Aufbau fachbezogener Kompetenzen genutzt wird (Leutner et al., 2004; OECD, 2004), kommt dieser Frage zudem eine besondere Bedeutung zu. So würden Hinweise auf die Bedeutung der analytischen Problemlösekompetenz für zukünftige fachbezogene Kompetenzen diese Hypothese weiter stützen.

Gegenstand von Studie 3 ist die Untersuchung von Aspekten der prognostischen Validität der Modellierung analytischer Problemlösekompetenz in Bezug auf die Vorhersage zukünftiger fachbezogener Kompetenzen. Auch in diesem Zusammenhang gilt es, analytische Problemlösekompetenz von ähnlichen und bewährten Konstrukten wie dem schlussfolgernden Denken (*reasoning*) beziehungsweise den fluiden Fähigkeiten als Kernkonzepte der Intelligenz (Carroll, 1993; Gustafsson, 1988), im Sinne der diskriminanten Validität, abzugrenzen. Dies kann darüber geschehen, dass ein Mehrwert der analytischen Problemlösekompetenz gegenüber fluiden Fähigkeiten bei der Vorhersage fachbezogener Kompetenzen gezeigt wird. Diese als *inkrementelle Validität* bezeichnete, eher angewandte Art der Validität (Sechrest, 1963; siehe auch Hunsley & Meyer, 2003; *Nützlichkeitsaspekt* als Bestandteil des *externen Aspekts* der Konstruktvalidität sensu Messick, 1995) ist daher ebenfalls Gegenstand von Studie 3. Hierzu werden Regressionsmodelle mit analytischer Problemlösekompetenz und fluiden Fähigkeiten als Prädiktoren zur Vorhersage zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften spezifiziert.

## 4.2 Theoretischer Hintergrund – Prädiktoren fachbezogener schulischer Kompetenzen

### 4.2.1 Allgemeine kognitive Grundfähigkeiten und fachbezogene Kompetenzen

Maße der Intelligenz gehören mit zu den erklärungskräftigsten Prädiktoren fachbezogener schulischer Kompetenzen (z. B. Rohde & Thompson, 2007; Rost, 2009, Kap. 6). Die Korrelationen liegen meist in einem Bereich zwischen  $r = .50$  bis  $r = .70$  (z. B. Mackintosh, 1998; Süß, 2001). Der Einfluss der Intelligenz auf die Leistungsentwicklung sowie den Schulerfolg wird in Längsschnittstudien deutlich. So konnte Kline (1991) zeigen, dass die Intelligenz, erfasst im Alter von fünf Jahren, den späteren Schulerfolg in verschiedenen Schulfächern vorhersagt (siehe auch Clemmer, Klifman & Bradley-Johnson, 1992; Leeson, Ciarrochi & Heaven, 2008). Ebenso konnten Deary, Strand, Smith und Fernandes (2007) nachweisen, dass die Intelligenz, erfasst im Alter von 11 Jahren, die Leistungen in einem standardisierten Abschlusstest im Alter von 16 Jahren gut vorhersagen konnte ( $r = .81$ ). Differenziert nach

Schulfächern zeigte sich für die Mathematik die höchste Korrelation ( $r = .77$ ). Die beschriebenen Ergebnisse lassen sich dahingehend interpretieren, dass intelligentere Schülerinnen und Schüler über effizientere Lern- und Problemlösestrategien verfügen und wissen, in welchen Situationen diese effektiv eingesetzt werden können. Sie sind dadurch in der Lage, sich schneller auf neue Aufgaben und Probleme einzustellen und lösungsrelevante Regeln zu erkennen (Helmke & Weinert, 1997).

#### 4.2.2 Vorwissen und fachbezogene Kompetenzen

Neben der Intelligenz ist auch das bereichsspezifische Vorwissen ein erklärungskräftiger Prädiktor für zukünftige fachbezogene Kompetenzen, dessen Bedeutung im Vergleich zu Maßen der Intelligenz im Verlauf der Schulkarriere sogar zuzunehmen scheint (Hany, 1997). So konnten Helmke und Weinert (1997) mit Bezug auf mathematische Kompetenzen zeigen, dass der Einfluss der Intelligenz über die Zeit abnimmt (von  $r = .30$  auf  $r = .14$ ; für ähnliche Ergebnisse in Bezug auf Indikatoren der Lesekompetenz siehe Ferrer et al., 2007) und gleichzeitig die mathematischen Kompetenzen zu einem früheren Zeitpunkt in ihrer Bedeutung als Prädiktoren für zukünftige mathematische Kompetenzen zunehmen (von  $r = .45$  auf  $r = .63$ ; vgl. Stern 2001a). Ebenso konnte Stern (2003) anhand von Daten der LOGIK-Follow-Up-Studie zeigen, dass die Leistungen im Lösen von mathematischen Textaufgaben in der 2. Klasse mit der Mathematikleistung in der 11. Klasse korrelierten, während die in der 2. Klasse erfasste Intelligenz keinerlei signifikanten Zusammenhang mit der Mathematikleistung in der 11. Klasse aufwies. Es ist jedoch nicht davon auszugehen, dass Vorwissen einen Ersatz für Intelligenz darstellt. In längsschnittlichen Untersuchungen hat sich in der Regel zum einen erwiesen, dass auch nach der Kontrolle des Vorwissens Intelligenzeffekte auf die Leistungsentwicklung zu finden sind. Zum anderen wird auch durch das Vorwissen über den Effekt der Intelligenz hinaus zusätzliche Varianz bezüglich der Schulleistungen aufgeklärt (z. B. Stern, 2001b). Beide besitzen also wechselseitig inkrementelle Validität.

Insgesamt lässt sich festhalten, dass sowohl Maße der Intelligenz als auch bereichsspezifisches Vorwissen relevante Prädiktoren für zukünftige fachbezogene Kompetenzen

darstellen. In Domänen, in denen Kompetenzen stärker das Produkt akkumulierter Lernprozesse sind, scheint die Bedeutung der Intelligenz im Laufe der Lerngeschichte geringer und die des bereichsspezifischen Vorwissens größer zu werden. Im Gegensatz dazu ist die Rolle der Intelligenz immer dann besonders groß, wenn lediglich geringe Vorkenntnisse vorhanden sind und weniger stark auf bereits Gelerntes aufgebaut werden kann (Köller & Baumert, 2008; siehe auch Ackerman & Beier, 2006).

#### **4.2.3 Analytische Problemlösekompetenz und fachbezogene Kompetenzen am Beispiel von PISA 2003**

Analytische Problemlösekompetenz weist, wie in Abschnitt 1.3.3 dargestellt, deutliche Ähnlichkeiten sowohl auf theoretisch-konzeptueller als auch auf empirischer Ebene zu fachbezogenen Kompetenzen auf, insbesondere in der Mathematik, aber auch in den Naturwissenschaften. Diese Ähnlichkeiten betreffen sowohl die im Rahmen der Bearbeitung von analytischen Problemlöseaufgaben und fachbezogenen Aufgaben zu durchlaufenden Prozessschritte als auch die im Ablauf des Problemlöseprozesses relevanten Komponenten und Anforderungen (Fleischer, Wirth, Rumann & Leutner, 2010; Rumann, Fleischer, Stawitz, Wirth und Leutner, 2010). Empirisch schlagen sich diese Ähnlichkeiten in vergleichsweise hohen (messfehlerbereinigten) Korrelationen zwischen diesen Kompetenzbereichen nieder. Für den Zusammenhang zwischen analytischer Problemlösekompetenz und Mathematikkompetenz in PISA 2003 zeigen sich (je nach zugrunde gelegter Stichprobe) Korrelationen von  $r = .89$  (OECD, 2005, Kap. 13) beziehungsweise  $r = .90$  (Leutner et al., 2004). Für den Zusammenhang zwischen analytischer Problemlösekompetenz und naturwissenschaftlicher Kompetenz in PISA 2003 zeigen sich (je nach verwendeter Stichprobe) Korrelationen von  $r = .78$  (OECD, 2005, Kap. 13) beziehungsweise  $r = .85$  (Leutner et al., 2004). Substanzielle Zusammenhänge sowohl mit mathematischer als auch mit naturwissenschaftlicher Kompetenz finden sich auch bei einer mehrdimensionalen Konzeption der analytischen Problemlösekompetenz, bei der verschiedene Arten von Problemstellungen differenziert werden, wie die in Studie 1 berichteten Korrelationen zeigen (siehe Abschnitt 2.5).

Die beschriebenen Ähnlichkeiten zwischen analytischer Problemlösekompetenz und fachbezogenen Kompetenzen konnten anhand querschnittlicher Studien gezeigt werden. Die Annahme liegt jedoch nahe, dass analytische Problemlösekompetenz nicht nur im Querschnitt mit fachbezogenen Kompetenzen in der Mathematik und in den Naturwissenschaften korreliert, sondern auch als Prädiktor für zukünftige fachbezogene Kompetenzen in diesen Domänen geeignet ist. Zudem weist die Modellierung analytischer Problemlösekompetenz, wie in Studie 2 dargestellt, trotz diskriminanter Validität gegenüber schlussfolgerndem Denken und fluiden Fähigkeiten, auch eine deutliche Nähe zu diesen Konstrukten auf. Schlussfolgerndes Denken und fluide Fähigkeiten, als Kernkonzepte der Intelligenz (Carroll, 1993; Gustafsson, 1988), gelten wiederum, wie zuvor dargestellt, als zentrale Prädiktoren für zukünftige fachbezogene Kompetenzen. Dies spricht ebenfalls dafür, dass analytische Problemlösekompetenz, ähnlich wie Maße der Intelligenz, einen erklärungskräftigen Prädiktor für fachbezogene Kompetenzen in der Mathematik und den Naturwissenschaften darstellt.

### 4.3 Fragestellungen und Hypothesen

*Prognostische Validität:* Wie zuvor dargestellt, weist die analytische Problemlösekompetenz sowohl zu fachbezogenen Kompetenzen, insbesondere in der Mathematik und den Naturwissenschaften, als auch zu den fluiden Fähigkeiten als einem erklärungskräftigen Prädiktor für fachbezogene Kompetenzen, eine große Nähe auf. Dies lässt die Vermutung zu, dass die analytische Problemlösekompetenz geeignet ist, zukünftige fachbezogene Kompetenzen in der Mathematik und den Naturwissenschaften vorherzusagen (Hypothese 1). Schwieriger zu beantworten ist die Frage, ob analytische Problemlösekompetenz vor dem Hintergrund der zentralen Bedeutung des bereichsspezifischen Vorwissens geeignet ist, zukünftige Kompetenzen in der Mathematik und den Naturwissenschaften auch unter Kontrolle der jeweiligen Ausgangskompetenzen vorherzusagen. Dies betrifft die Frage der Vorhersagbarkeit der Kompetenzentwicklung in der Mathematik und den Naturwissenschaften durch die analytische Problemlösekompetenz (Forschungsfrage 1a). Weder die theoretischen Grundlagen noch bisherige empirische Ergebnisse erscheinen bezüglich dieser Fragestellung hinreichend tragfähig, sodass auf die Formulierung einer entsprechenden

Forschungshypothese verzichtet wird. Gleiches gilt für die Frage, ob analytische Problemlösekompetenz die Kompetenzentwicklung in der Mathematik und den Naturwissenschaften auch unter Kontrolle der fluiden Fähigkeiten vorhersagen kann (Forschungsfrage 1b).

*Inkrementelle Validität:* Darüber hinaus stellt sich die Frage, ob analytische Problemlösekompetenz auch über bereichsspezifisches Vorwissen und Intelligenzmaße hinaus, inkrementelle Varianz zukünftiger fachbezogener Kompetenzen aufklärt. Für ein Inkrement gegenüber fluiden Fähigkeiten spricht, dass Leutner et al. (2004) für eine Teilstichprobe aus PISA 2003 zeigen konnten, dass die analytische Problemlösekompetenz und fachbezogene Kompetenzen in der Mathematik und den Naturwissenschaften deutlich über den Anteil hinaus korrelieren, der durch die fluiden Fähigkeiten erklärbar ist. So verringerten sich die jeweiligen Korrelationen nach Auspartialisierung der fluiden Fähigkeiten nur relativ geringfügig (für Mathematik von  $r = .90$  auf  $r = .78$ ; für Naturwissenschaften von  $r = .85$  auf  $r = .71$ ; Leutner et al., 2004, S. 167). In dieselbe Richtung deuten auch die Ergebnisse des Nested-Faktor-Modells aus Studie 2 der vorliegenden Arbeit. Hier zeigten sich für den um fluide Fähigkeiten residualisierten Faktor der analytischen Problemlösekompetenz  $PI'$  im Querschnitt unter anderem substantielle Korrelationen mit fachbezogenen Kompetenzen in der Mathematik und den Naturwissenschaften (siehe Abschnitt 3.5). Es wird daher davon ausgegangen, dass die analytische Problemlösekompetenz, über den Effekt der fluiden Fähigkeiten hinaus, inkrementelle Varianz zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften aufklärt (Hypothese 2). Analog zur Forschungsfrage 1a wird, im Kontext der inkrementellen Validität, die Frage untersucht, ob analytische Problemlösekompetenz, über die Effekte der jeweiligen Ausgangsgangskompetenzen hinaus, inkrementelle Varianz zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften aufklärt (Forschungsfrage 2a). Ebenso wird, analog zur Forschungsfrage 1b, der Frage nachgegangen, ob analytische Problemlösekompetenz, auch über den Effekt der fluiden Fähigkeiten hinaus, inkrementelle Varianz der Kompetenzentwicklung in der Mathematik und in den Naturwissenschaften aufklärt (Forschungsfrage 2b).



## 4.4 Methode

### 4.4.1 Stichprobe und Datengrundlage

Die Untersuchung basiert auf den Daten der PISA-I-Plus Studie, einer nationalen Erweiterungsstudie zu PISA 2003 in Deutschland (PISA-Konsortium Deutschland, 2006b). Im Rahmen von PISA-I-Plus wurden  $N = 6020$  Schülerinnen und Schüler (43.9 % weiblich) in Klassenstufe 9 sowie ein Jahr später in Klassenstufe 10 erneut getestet. Die Stichprobe setzt sich aus 152 Schulen der Schulformen Realschule ( $n = 2199$ ), Schule mit mehreren Bildungsgängen ( $n = 653$ ), Integrierte Gesamtschule ( $n = 504$ ) und Gymnasium ( $n = 2664$ ) zusammen. Für die Kompetenzbereiche Mathematik und Naturwissenschaften liegen in dieser Studie Daten von zwei Messzeitpunkten vor, während Problemlösekompetenz und fluide Fähigkeiten nur zum ersten Messzeitpunkt in Klassenstufe 9 erhoben wurden. Für die detaillierte Beschreibung der Anlage sowie der methodischen Grundlagen von PISA-I-Plus sei auf Prenzel, Carstensen, Schöps und Maurischat (2006) sowie Carstensen (2006) verwiesen.

### 4.4.2 Instrumente

Zur Erfassung der Mathematikkompetenz, der naturwissenschaftlichen Kompetenz (jeweils in Klasse 9 und Klasse 10) sowie der analytischen Problemlösekompetenz (Klasse 9) wurden die Daten der entsprechenden PISA-Testskalen<sup>7</sup> herangezogen (PISA-Konsortium Deutschland, 2006a, Teil I, Kap. 1, 2 und 5). Zur Erfassung der fluiden Fähigkeiten (Klasse 9) wurden die Daten der in PISA eingesetzten Subskala *N2 Figurenanalogien* des *Kognitiven Fähigkeitstests* (KFT 5-12+R, Heller & Perleth, 2000; siehe PISA-Konsortium Deutschland, 2006a, Teil I, Kap. 6) verwendet. Diese Subskala erfasst schlussfolgerndes Denken (*reasoning*), welches als guter Indikator für fluide Fähigkeiten und ebenso für allgemeine kognitive Fähigkeiten gilt (Horn & Noll, 1997; Rost, 2009, Kap. 2; siehe hierzu auch die Ausführungen in Abschnitt 3.2.1).

---

<sup>7</sup> Für die Mathematikkompetenz in Klasse 10 wurden die Daten des curricular orientierten Tests verwendet, da für diesen Test die größten Veränderungen zwischen beiden Messzeitpunkten zu erwarten sind (Prenzel et al., 2006).

#### 4.4.3 Vorgehen und statistische Analysen

Alle Analysen erfolgten auf Basis der jeweils fünf *Plausible Values* der zuvor beschriebenen Testskalen. Bei Plausible Values handelt es sich um Zufallsziehungen aus den Wahrscheinlichkeitsverteilungen der Kompetenzen für jede Person, zu deren Schätzung neben den jeweiligen Testdaten auch zusätzliche Hintergrundinformationen über die Schülerinnen und Schüler herangezogen werden und die dadurch unter anderem eine hohe Messgenauigkeit besitzen (Carstensen, Knoll, Rost & Prenzel, 2004; OECD, 2005, Kap. 9). Zur Prüfung der prognostischen und inkrementellen Validität der Modellierung analytischer Problemlösekompetenz wurden für die Mathematikkompetenz und die naturwissenschaftliche Kompetenz (jeweils in Klasse 10) als Outcome-Variablen jeweils getrennte Regressionsanalysen durchgeführt. In diese Analysen wurden die analytische Problemlösekompetenz, die fluiden Fähigkeiten und die jeweiligen Ausgangskompetenzen in Mathematik beziehungsweise Naturwissenschaften (jeweils Klasse 9) als Prädiktor-Variablen einbezogen. Zur Untersuchung der Kompetenzentwicklung wurde in den Regressionsanalysen für die Mathematikkompetenz beziehungsweise die naturwissenschaftliche Kompetenz (jeweils in Klasse 9) kontrolliert. Zur Bewertung des inkrementellen Wertes einzelner Prädiktoren wird, den Empfehlungen von Hunsley und Meyer (2003) folgend, bei Hinzunahme eines zweiten Prädiktors in die Regressionsgleichung ein Zuwachs der aufgeklärten Varianz ( $R^2$ ) um  $\Delta R^2 = .09$  und bei Hinzunahme eines dritten Prädiktors ein Zuwachs um  $\Delta R^2 = .0225$  als substanzielles Inkrement gewertet. Diesen Benchmarks liegt die Rationale zugrunde, dass ein Prädiktor mindestens einen kleinen bis mittleren Effekt nach Cohen (1992) haben sollte, was einer semipartiellen Korrelation von  $r = .30$  und dies wiederum einem  $\Delta R^2$  von  $.09$  entspricht. Basierend auf der Beobachtung, dass Zuwächse im  $R^2$  bei Hinzunahme eines dritten Prädiktors generell sehr gering ausfallen, schlagen Hunsley und Meyer (2003) für den dritten Prädiktor als Untergrenze eine semipartielle Korrelation von  $r = .15$  vor, was einem  $\Delta R^2$  von  $.0225$  entspricht. Um die in PISA verwendete Plausible-Value-Technik (OECD, 2005) sowie die geclusterte Datenstruktur (Schülerinnen und Schüler geclustert in Schulen) simultan berücksichtigen zu können, erfolgten alle Analysen mit dem Programm WesVar <sup>8</sup>

---

<sup>8</sup> Da nur für den PISA-I-Plus-Datensatz die für die Berechnungen in WesVar notwendigen Replikationsgewichte vorlagen, konnte dieses Programm nur in Studie 3 verwendet werden, während in den Studien 1, 2 und 4 das Programm HLM 7 eingesetzt wurde.

(Westat, 2012; zum Umgang mit Plausible Values und geclusterten Datenstrukturen siehe die Ausführungen in Studie 1 in Abschnitt 2.4.3).

## 4.5 Ergebnisse

### 4.5.1 Vorhersage der Mathematikkompetenz in Klasse 10

Tabelle 4.1 stellt die Ergebnisse zur Vorhersage der Mathematikkompetenz in Klasse 10 durch die in Klasse 9 erhobenen Prädiktoren fluide Fähigkeiten, Mathematikkompetenz und analytische Problemlösekompetenz dar. Die Prädiktorvariablen korrelierten untereinander wie folgt: fluide Fähigkeiten und Mathematik:  $r = .65$ ; fluide Fähigkeiten und Problemlösen:  $r = .67$ ; Mathematik und Problemlösen:  $r = .74$ . Wie im Vergleich der Modelle 1 bis 3 deutlich wird, hatte die bereits in Klasse 9 vorhandene Mathematikkompetenz mit  $\beta = .845$  den höchsten Effekt auf die Mathematikkompetenz in Klasse 10. Ebenso leistete die analytische Problemlösekompetenz in Klasse 9 mit  $\beta = .701$  einen signifikanten Beitrag zur Vorhersage der Mathematikkompetenz in Klasse 10, der zudem signifikant größer war als der Beitrag der fluiden Fähigkeiten in Klasse 9 ( $\beta = .638$ ), wie der Vergleich der entsprechenden 95%-Konfidenzintervalle zeigt ( $KI_{\text{fluide Fähigkeiten}}: .607/.669$ ;  $KI_{\text{Problemlösen}}: .674/.728$ ). Die analytische Problemlösekompetenz stellt demnach, im Sinne der prognostischen Validität, einen erklärungskräftigen Prädiktor für zukünftige Mathematikkompetenz dar (Hypothese 1). Auch bei gleichzeitiger Kontrolle der fluiden Fähigkeiten zeigte sich ein signifikanter Effekt der analytischen Problemlösekompetenz auf die Mathematikkompetenz in Klasse 10, wie Modell 4 verdeutlicht ( $\beta = .497$ ). Auch in diesem Modell erwies sich die analytische Problemlösekompetenz als signifikant stärkerer Prädiktor im Vergleich zu den fluiden Fähigkeiten ( $KI_{\text{fluide Fähigkeiten}}: .264/.334$ ;  $KI_{\text{Problemlösen}}: .462/.532$ ). Ebenso leistete die analytische Problemlösekompetenz auch unter Kontrolle der Mathematikkompetenz in Klasse 9 einen signifikanten Beitrag zur Vorhersage der Mathematikkompetenz in Klasse 10 ( $\beta = .166$ ), wie Modell 5 zeigt. Gleiches galt auch unter zusätzlicher Kontrolle der fluiden Fähigkeiten, wie in Modell 7 ersichtlich ist. Die analytische Problemlösekompetenz leistete somit einen signifikanten Beitrag zur Vorhersage der Kompetenzentwicklung in der Mathematik (Forschungsfrage 1a) auch unter Kontrolle der fluiden Fähigkeiten (Forschungsfrage 1b).

Der Vergleich der Modelle 1 und 4 zeigt, dass durch die Hinzunahme der analytischen Problemlösekompetenz als weiteren Prädiktor gegenüber den fluiden Fähigkeiten als alleinigem Prädiktor die aufgeklärte Varianz von  $R^2 = .406$  auf  $R^2 = .539$  zunahm. Diese Veränderung von  $\Delta R^2 = .133$  gilt nach dem von Hunsley und Meyer (2003) vorgeschlagenen Kriterium von  $\Delta R^2 = .09$  als substantiell. Die analytische Problemlösekompetenz klärte somit, im Sinne der inkrementellen Validität, über die fluiden Fähigkeiten hinaus, substantiell zusätzliche Varianz bezüglich der Mathematikkompetenz in Klasse 10 auf (Hypothese 2). Der Vergleich der Modelle 3 und 4 verdeutlicht, dass umgekehrt die fluiden Fähigkeiten keine substantiell inkrementelle Varianz gegenüber der analytischen Problemlösekompetenz aufklärten. Der Zuwachs an aufgeklärter Varianz betrug hier lediglich  $\Delta R^2 = .048$ .

Die Hinzunahme der analytischen Problemlösekompetenz als weiteren Prädiktor in ein Modell mit der Ausgangskompetenz in Mathematik als alleinigem Prädiktor (Modelle 2 und 5) führte zu einer Zunahme der aufgeklärten Varianz um  $\Delta R^2 = .020$ . Dies verfehlt die von Hunsley und Meyer (2003) vorgeschlagene Grenze für ein substantielles Inkrement deutlich (Forschungsfrage 2a). Gleiches gilt für die fluiden Fähigkeiten, wie der Vergleich der Modelle 2 und 6 deutlich macht. So führte auch die Hinzunahme der fluiden Fähigkeiten in ein Modell mit der Ausgangskompetenz in Mathematik als alleinigem Prädiktor nur zu einer nicht substantiellen Zunahme der aufgeklärten Varianz ( $\Delta R^2 = .020$ ).

Wie ein Vergleich der Modelle 6 und 7 zeigt, führte die zusätzliche Berücksichtigung der analytischen Problemlösekompetenz, gegenüber einem Modell mit den fluiden Fähigkeiten und der Ausgangskompetenz in Mathematik als Prädiktoren, nur zu einer nicht substantiellen Zunahme der aufgeklärten Varianz von  $\Delta R^2 = .005$ . Die analytische Problemlösekompetenz klärte somit, über den Effekt der fluiden Fähigkeiten hinaus, keine substantiell inkrementelle Varianz der Kompetenzentwicklung in der Mathematik auf (Forschungsfrage 2b). Umgekehrt klärten auch die fluiden Fähigkeiten, über den Effekt der analytischen Problemlösekompetenz in Klasse 9 hinaus, keine substantiell inkrementelle Varianz der Kompetenzentwicklung in der Mathematik auf, wie der Vergleich der Modelle 5 und 7 verdeutlicht ( $\Delta R^2 = .005$ ).

**Tabelle 4.1: Regressionsanalysen zur Vorhersage der Mathematikkompetenz (Klasse 10) durch fluide Fähigkeiten, Mathematikkompetenz und analytische Problemlösekompetenz (jeweils Klasse 9): Standardisierte Regressionskoeffizienten ( $\beta$ ; 95%-Konfidenzintervalle [KI] in eckigen Klammern) und aufgeklärte Varianz ( $R^2$ )**

Mathematikkompetenz (curricular orientierter Test; Klasse 10)							
Prädiktoren (Klasse 9)	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5	Modell 6	Modell 7
	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]
Fluide Fähigkeiten	.638 [.607/.669]			.299 [.264/.334]		.149 [.120/.178]	.108 [.079/.137]
Mathematikkompetenz		.845 [.831/.859]			.721 [.688/.754]	.747 [.722/.772]	.685 [.648/.722]
Analytische Problemlösekompetenz			.701 [.674/.728]	.497 [.462/.532]	.166 [.131/.201]		.119 [.084/.154]
$R^2$	.406	.706	.491	.539	.726	.726	.731

Anmerkungen. Für alle Regressionskoeffizienten gilt:  $p < .001$ ;  $N = 6020$ .

#### 4.5.2 Vorhersage der naturwissenschaftlichen Kompetenz in Klasse 10

Tabelle 4.2 stellt die Ergebnisse zur Vorhersage der naturwissenschaftlichen Kompetenz in Klasse 10 durch die Prädiktoren fluide Fähigkeiten, naturwissenschaftliche Ausgangskompetenz und analytische Problemlösekompetenz jeweils in Klasse 9 dar. Die Prädiktoren korrelierten untereinander wie folgt: fluide Fähigkeiten und Naturwissenschaften:  $r = .61$ ; fluide Fähigkeiten und Problemlösen:  $r = .67$ ; Naturwissenschaften und Problemlösen:  $r = .71$ . Auch bei der Vorhersage der naturwissenschaftlichen Kompetenz zeigte sich, wie bereits bei der Vorhersage der Mathematikkompetenz in Klasse 10 (siehe Tabelle 4.1), der größte Effekt für die entsprechende Ausgangskompetenz, in diesem Fall in den Naturwissenschaften ( $\beta = .781$ ). Ebenso leistete auch die analytische Problemlösekompetenz in Klasse 9 mit  $\beta = .661$  einen signifikanten Beitrag zur Vorhersage der naturwissenschaftlichen Kompetenz in Klasse 10, der zudem signifikant höher ausfiel als der Beitrag der fluiden Fähigkeiten ( $\beta = .575$ ), wie ein Vergleich der entsprechenden 95%-Konfidenzintervalle deutlich macht. Die analytische Problemlösekompetenz stellt demnach, im Sinne der prognostischen Validität, einen erklärungskräftigen Prädiktor für zukünftige naturwissenschaftliche Kompetenzen dar (Hypothese 1). Auch bei gleichzeitiger Kontrolle der fluiden Fähigkeiten zeigte sich ein signifikanter Effekt der analytischen Problemlösekompetenz auf die naturwissenschaftliche Kompetenz in Klasse 10, wie Modell 4 verdeutlicht ( $\beta = .502$ ). Auch in diesem Modell erwies sich, wie bereits im entsprechenden Modell bei der Vorhersage der Mathematikkompetenz, die analytische Problemlösekompetenz als signifikant stärkerer Prädiktor im Vergleich zu fluiden Fähigkeiten ( $KI_{\text{fluide Fähigkeiten}}: .190/.276$ ;  $KI_{\text{Problemlösen}}: .461/.543$ ). Ebenso leistete die analytische Problemlösekompetenz auch unter Kontrolle der naturwissenschaftlichen Kompetenz in Klasse 9 einen signifikanten Beitrag zur Vorhersage der naturwissenschaftlichen Kompetenz in Klasse 10 ( $\beta = .215$ ), wie Modell 5 zeigt. Gleiches galt auch unter zusätzlicher Kontrolle der fluiden Fähigkeiten, wie in Modell 7 ersichtlich ist. Die analytische Problemlösekompetenz leistete somit einen signifikanten Beitrag zur Vorhersage der Kompetenzentwicklung in den Naturwissenschaften (Forschungsfrage 1a), auch unter Kontrolle der fluiden Fähigkeiten (Forschungsfrage 1b).

Durch die Hinzunahme der analytischen Problemlösekompetenz als weiteren Prädiktor in Modell 4 wurde die aufgeklärte Varianz gegenüber Modell 1 mit den fluiden Fähig-

keiten als alleinigem Prädiktor substanziell um  $\Delta R^2 = .135$  erhöht. Die analytische Problemlösekompetenz klärte somit, im Sinne der inkrementellen Validität, über die fluiden Fähigkeiten hinaus, zusätzliche Varianz bezüglich der naturwissenschaftlichen Kompetenz in Klasse 10 auf (Hypothese 2). Umgekehrt klärten die fluiden Fähigkeiten, wie bereits bei der Vorhersage der Mathematikkompetenz in Klasse 10, auch bei der Vorhersage der naturwissenschaftlichen Kompetenz in Klasse 10, keine substanziell inkrementelle Varianz gegenüber der analytischen Problemlösekompetenz auf, wie der Vergleich der Modelle 3 und 4 verdeutlicht ( $\Delta R^2 = .029$ ).

Die Berücksichtigung der analytischen Problemlösekompetenz als zusätzlicher Prädiktor in Modell 5, gegenüber der Ausgangskompetenz in den Naturwissenschaften als alleinigem Prädiktor in Modell 2, führte lediglich zu einer nicht substanziellen Zunahme der aufgeklärten Varianz um  $\Delta R^2 = .023$  (Forschungsfrage 2a). Dies gilt ebenso für die fluiden Fähigkeiten, wie im Vergleich der Modelle 2 und 6 deutlich wird. So führte auch die Hinzunahme der fluiden Fähigkeiten in ein Modell mit der naturwissenschaftlichen Ausgangskompetenz als alleinigem Prädiktor nur zu einer nicht substanziellen Zunahme der aufgeklärten Varianz ( $\Delta R^2 = .015$ ).

Die zusätzliche Berücksichtigung der analytischen Problemlösekompetenz in Modell 7 gegenüber Modell 6 mit den fluiden Fähigkeiten und der Ausgangskompetenz in den Naturwissenschaften als Prädiktoren, führte lediglich zu einer nicht substanziellen Zunahme der aufgeklärten Varianz von  $\Delta R^2 = .012$ . Die analytische Problemlösekompetenz klärte somit, über den Effekt der fluiden Fähigkeiten hinaus, keine substanziell inkrementelle Varianz der Kompetenzentwicklung in den Naturwissenschaften auf (Forschungsfrage 2b). Umgekehrt klärten auch die fluiden Fähigkeiten, über den Effekt der analytischen Problemlösekompetenz in Klasse 9 hinaus, keine substanziell inkrementelle Varianz der Kompetenzentwicklung in den Naturwissenschaften auf, wie der Vergleich der Modelle 5 und 7 zeigt ( $\Delta R^2 = .004$ ).

**Tabelle 4.2: Regressionsanalysen zur Vorhersage der naturwissenschaftlichen Kompetenz (Klasse 10) durch fluide Fähigkeiten, naturwissenschaftliche Kompetenz und analytische Problemlösekompetenz (jeweils Klasse 9): Standardisierte Regressionskoeffizienten ( $\beta$ ; 95%-Konfidenzintervalle [KI] in eckigen Klammern) und aufgeklärte Varianz ( $R^2$ )**

Prädiktoren (Klasse 9)	Naturwissenschaftliche Kompetenz (Klasse 10)						
	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5	Modell 6	Modell 7
	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]	$\beta$ [KI]
Fluide Fähigkeiten	.575 [.542/.608]			.233 [.190/.276]		.154 [.119/.189]	.086 [.047/.125]
Naturwissenschaftliche Kompetenz		.781 [.763/.799]			.628 [.599/.657]	.686 [.661/.711]	.605 [.576/.634]
Analytische Problemlösekompetenz			.661 [.636/.686]	.502 [.461/.543]	.215 [.184/.246]		.173 [.136/.210]
$R^2$	.331	.610	.437	.466	.633	.625	.637

*Anmerkungen.* Für alle Regressionskoeffizienten gilt:  $p < .001$ ;  $N = 6020$ .



## 4.6 Diskussion

### 4.6.1 Zusammenfassung der Ergebnisse

Gegenstand von Studie 3 war die Untersuchung von Aspekten der prognostischen und der inkrementellen Validität der Modellierung analytischer Problemlösekompetenz im Vergleich zu fluiden Fähigkeiten bei der Vorhersage zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften. Hierzu wurde mithilfe von Regressionsanalysen der Einfluss der analytischen Problemlösekompetenz in Klasse 9 auf fachbezogene Kompetenzen in der Mathematik und den Naturwissenschaften in Klasse 10 untersucht. Hypothese 1 konnte dabei bestätigt werden. So ist die analytische Problemlösekompetenz geeignet, zukünftige fachbezogene Kompetenzen in der Mathematik und in den Naturwissenschaften vorherzusagen. Dies gilt auch für die Vorhersage der Kompetenzentwicklung in der Mathematik und in den Naturwissenschaften (Forschungsfrage 1a) selbst unter Kontrolle der fluiden Fähigkeiten (Forschungsfrage 1b).

Ebenso konnte Hypothese 2 bestätigt werden. Die analytische Problemlösekompetenz klärt, über den Effekt der fluiden Fähigkeiten hinaus, substantiell inkrementelle Varianz zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften auf.<sup>9</sup> Unter Kontrolle der jeweiligen Ausgangskompetenzen in der Mathematik und den Naturwissenschaften zeigt sich jedoch keine substantiell inkrementelle Varianzaufklärung zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften durch die analytische Problemlösekompetenz (Forschungsfrage 2a). Ebenso wenig zeigt sich bei der Vorhersage der Kompetenzentwicklung ein substantielles Inkrement der analytischen Problemlösekompetenz gegenüber den fluiden Fähigkeiten (Forschungsfrage 2b). Gleiches gilt umgekehrt für die fluiden Fähigkeiten, auch hier zeigt sich weder bei der Vorhersage zukünftiger Kompetenzen in der Mathematik und den Naturwissenschaften, noch bei der Vorhersage der Kompetenzentwicklung, ein substantielles Inkrement gegenüber der analytischen Problemlösekompetenz.

---

<sup>9</sup> Interessanterweise gilt dies umgekehrt nicht für die fluiden Fähigkeiten, für die sich kein substantielles Inkrement gegenüber der analytischen Problemlösekompetenz zeigen lässt.

#### 4.6.2 Theoretische und praktische Implikationen

Die in den Studien 1 und 2 gewonnenen Erkenntnisse zur faktoriellen und diskriminanten Validität der Modellierung analytischer Problemlösekompetenz konnten in Studie 3 um Ergebnisse zur prognostischen und inkrementellen Validität erweitert werden. Querschnittliche Zusammenhänge der analytischen Problemlösekompetenz mit fachbezogenen Kompetenzen, insbesondere in der Mathematik und den Naturwissenschaften, im Sinne der Übereinstimmungsvalidität (als Aspekt der Kriteriumsvalidität, z. B. Hartig, Frey & Jude, 2012; oder als externer Aspekt der Konstruktvalidität, sensu Messick, 1995), sind, wie zuvor dargestellt, vielfach belegt (z. B. Leutner et al., 2004; OECD, 2005). Die vorliegende Untersuchung erweitert den Kenntnisstand dahingehend, dass auch in längsschnittlicher Perspektive Zusammenhänge mit fachbezogenen Kompetenzen in der Mathematik und den Naturwissenschaften im Sinne der prognostischen Validität gezeigt werden konnten. Die Ergebnisse weisen damit auf die prinzipielle Bedeutung der analytischen Problemlösekompetenz hin – und zwar, sowohl für die Vorhersage *zukünftiger* fachbezogener Kompetenzen als auch für die Vorhersage der *Kompetenzentwicklung*. Letzteres ist vor dem Hintergrund der zentralen Bedeutung des bereichsspezifischen Vorwissens besonders bemerkenswert, zumal dieser Effekt auch unter zusätzlicher Kontrolle der fluiden Fähigkeiten als zentralem Intelligenzkonzept (Carroll, 1993; Gustafsson, 1988; Horn & Noll, 1997; Mackintosh, 1998; Rost, 2009, Kap. 2) bestehen bleibt. Die analytische Problemlösekompetenz erweist sich bei der Vorhersage zukünftiger fachbezogener Kompetenzen im Vergleich zu schlussfolgerndem Denken beziehungsweise fluiden Fähigkeiten sogar als der „bessere“, weil erklärungskräftigere Prädiktor. Dies gilt sowohl für die Betrachtung der Effekte beider Variablen, der analytischen Problemlösekompetenz und der fluiden Fähigkeiten, als Prädiktoren in getrennten Modellen (siehe Modell 1 versus Modell 3 in Tabelle 4.1 und 4.2) als auch bei wechselseitiger Kontrolle innerhalb eines Modells (siehe Modell 4 in Tabelle 4.1 und 4.2). Die mit dem analytischen Problemlösetest erfassten Fähigkeiten scheinen somit für die Bearbeitung zukünftiger mathematischer und naturwissenschaftlicher Aufgabenstellungen in stärkerem Maße lösungsrelevant zu sein, als diejenigen, die durch den eingesetzten Test für fluide Fähigkeiten erfassten werden. Bei der Vorhersage der Kompetenzentwicklung

zeigt sich diese stärkere prädiktive Kraft der analytischen Problemlösekompetenz zumindest eingeschränkt, bei wechselseitiger Kontrolle innerhalb eines Modells, für die naturwissenschaftlichen Kompetenzen (siehe Modell 7 in Tabelle 4.2).

Neben Hinweisen zur prognostischen Validität konnte in der vorliegenden Untersuchung auch ein substanzielles Inkrement der analytischen Problemlösekompetenz gegenüber fluiden Fähigkeiten bei der Vorhersage sowohl der mathematischen als auch der naturwissenschaftlichen Kompetenzen gezeigt werden. Umgekehrt konnte für die fluiden Fähigkeiten jedoch keine derartige substanzielle Zunahme der aufgeklärten Varianz bei der Vorhersage der beiden fachbezogenen Kompetenzen im Sinne der inkrementellen Validität gezeigt werden. Die Ergebnisse bestätigen damit die Befunde aus Studie 2, in der für einen von fluiden Fähigkeiten unabhängigen Faktor der analytischen Problemlösekompetenz substanzielle Korrelationen mit fachbezogenen Kompetenzen unter anderem in der Mathematik und den Naturwissenschaften gezeigt werden konnten. Dies verdeutlichte bereits den Mehrwert der analytischen Problemlösekompetenz bei der Vorhersage fachbezogener Kompetenzen, also ihre inkrementelle Varianzaufklärung, gegenüber fluiden Fähigkeiten, allerdings in einem querschnittlichen Design. Die hier dargestellten Ergebnisse aus Studie 3 stehen darüber hinaus in Einklang mit Fischer, Greiff, Wüstenberg, Fleischer, Buchwald und Funke (2015), die, allerdings ebenfalls in einer querschnittlichen Untersuchung, sowohl für Schülerinnen und Schüler als auch für Studierende inkrementelle Varianzaufklärung der analytischen Problemlösekompetenz bei der Vorhersage von Schulnoten über die Effekte von fluiden Fähigkeiten und komplexer Problemlösekompetenz hinaus zeigen konnten. Die vorliegende Untersuchung erweitert dieses Befundmuster, indem ein entsprechendes Inkrement der analytischen Problemlösekompetenz gegenüber fluiden Fähigkeiten auch in längsschnittlicher Perspektive belegt werden konnte. Dieser Befund weist erneut auf die Bedeutung und Nützlichkeit der analytischen Problemlösekompetenz für die Vorhersage zukünftiger fachbezogener Kompetenzen hin und liefert zudem ein weiteres Argument für eine empirische Trennung (diskriminante Validität) von analytischer Problemlösekompetenz und fluiden Fähigkeiten, und zwar nicht nur im Querschnitt, sondern auch in längsschnittlichen Untersuchungen (für Befunde zur inkrementellen Validität von Maßen zur Erfassung dynamischer bzw. komplexer Problemlösekompetenz siehe z. B. Greiff et al., 2013; Sonnleitner, Keller, Martin & Brunner, 2013; Wüstenberg, Greiff & Funke, 2012).

Die Tatsache, dass kein substanzielles Inkrement der analytischen Problemlösekompetenz gegenüber den jeweiligen Ausgangskompetenzen bei der Vorhersage zukünftiger fachbezogener Kompetenzen und somit auch kein substanzielles Inkrement gegenüber den fluiden Fähigkeiten bei der Vorhersage der Kompetenzentwicklung gezeigt werden konnte, unterstreicht einmal mehr die zentrale Bedeutung des bereichsspezifischen Vorwissens für zukünftige fachbezogene Kompetenzen quasi zu lasten der analytischen Problemlösekompetenz und der fluiden Fähigkeiten. Es ist jedoch davon auszugehen, dass sich ein nicht unerheblicher Teil des Einflusses der analytischen Problemlösekompetenz auf die fachbezogenen Kompetenzen bereits zum ersten Messzeitpunkt manifestiert hat und somit im Sinne eines indirekten Effekts über die jeweiligen Ausgangskompetenzen auf die fachbezogenen Kompetenzen in Klasse 10 vermittelt wird. Für diese angenommene Mediation spricht die signifikante Reduktion der Regressionskoeffizienten der analytischen Problemlösekompetenz bei Kontrolle der jeweiligen Ausgangskompetenzen (Modell 3 versus Modell 5 in Tabelle 4.1 und 4.2; Sobel-Test<sub>Mathematik</sub>:  $t = 34.91$ ;  $SE = .015$ ;  $p < .001$ ; Sobel-Test<sub>Naturwissenschaften</sub>:  $t = 31.42$ ;  $SE = .014$ ;  $p < .001$ ; siehe MacKinnon, Lockwood, Hoffman, West & Sheets, 2002). Es ist somit davon auszugehen, dass die Bedeutung der analytischen Problemlösekompetenz für die Entwicklung von fachbezogenen Kompetenzen in den präsentierten Analysen noch unterschätzt wird. Dies lässt sich jedoch auf Basis der verwendeten Datengrundlage nicht abschließend prüfen. Festzuhalten ist außerdem, dass das mangelnde Inkrement gegenüber den jeweiligen Ausgangskompetenzen sowie gegenüber den fluiden Fähigkeiten bei der Vorhersage der Kompetenzentwicklung kein Spezifikum der analytischen Problemlösekompetenz darstellt. Es trifft umgekehrt für die fluiden Fähigkeiten gegenüber der analytischen Problemlösekompetenz in gleichem Maße zu.

Die vorliegenden Ergebnisse zur prognostischen und inkrementellen Validität der Modellierung analytischer Problemlösekompetenz deuten darauf hin, dass die analytische Problemlösekompetenz beim Erwerb fachbezogener Kompetenzen eine wichtige Rolle spielen kann. Inwiefern die analytische Problemlösekompetenz eine kognitive Ressource darstellt und auf welche Weise sie sich – wie die Potenzialausschöpfungshypothese postuliert – didaktisch zum Aufbau fachbezogener Kompetenzen in der Praxis nutzen lässt, muss im Rahmen zukünftiger Interventionsstudien geklärt werden. Dies ist insofern von besonderem Interesse, als dass sich durch die Förderung einer einzelnen fächerübergreifenden

Kompetenz Transfermöglichkeiten in verschiedene Fachdomänen, insbesondere in die Mathematik und die Naturwissenschaften, ergeben würden (Fleischer et al., 2010; Rumann et al., 2010). Für die Mathematik liegen hierzu bereits erste Hinweise vor. So untersuchten Buchwald, Fleischer und Leutner (2015) in einem Prä-Post-Design Transfereffekte eines mehrwöchigen Problemlösetrainings – mit Schwerpunkt auf konditionalem und prozeduralem Wissen sowie Planungsfähigkeit und metakognitiven Heuristiken als gemeinsame Komponenten von Problemlösekompetenz und mathematischer Kompetenz (Fleischer et al., 2010; siehe hierzu auch die Ausführungen in Abschnitt 1.3.2) – bei Schülerinnen und Schülern der 9. Jahrgangsstufe. Die Autoren konnten dabei einen kleinen, aber substantiellen Transfereffekt auf die Mathematikkompetenz bei schwachen Problemlösern zeigen (Buchwald, Fleischer & Leutner, 2015; siehe auch Buchwald, Fleischer, Rumann, Wirth & Leutner, 2017). Dieser Befund steht im Einklang mit den nationalen Befunden aus PISA 2003, die kognitive Potenziale insbesondere bei Schülerinnen und Schülern im unteren Leistungsbereich erkennen lassen (Leutner et al., 2004; Leutner, Klieme, Meyer & Wirth, 2005).

#### 4.6.3 Grenzen und Ausblick

Die dargestellten Ergebnisse von Studie 3 basieren auf Daten von zwei Messzeitpunkten, wobei die analytische Problemlösekompetenz lediglich zum ersten Messzeitpunkt erhoben wurde. Die Datengrundlage erlaubt damit keine kausale Prüfung der angesprochenen Mediationseffekte der analytischen Problemlösekompetenz auf die fachbezogenen Mathematik- und Naturwissenschaftskompetenzen in Klasse 10 über die jeweiligen Ausgangskompetenzen in Klasse 9. Daher wurden keine Strukturgleichungsmodelle oder Pfadanalysen mit entsprechenden indirekten Effekten spezifiziert, sondern Regressionsmodelle zur Prüfung der Hypothesen verwendet, zumal die Prüfung von Mediationseffekten nicht Gegenstand von Studie 3 war.

Um den Einfluss der analytischen Problemlösekompetenz auf zukünftige fachbezogene Kompetenzen und auch auf die Entwicklung dieser Kompetenzen besser abschätzen zu können, müssen zukünftige Studien entsprechende Datenstrukturen bereitstellen. Eine bessere Grundlage für Kausalinterpretationen würde beispielsweise ein Cross-Lagged-

Panel-Design über drei Messzeitpunkte liefern (z. B. McArdle, 2009). Dies würde zum einen die Schätzung entsprechender Cross-Lagged-Effekte im Rahmen autoregressiver Modelle erlauben, und damit sowohl die Prüfung der angesprochenen Mediationseffekte, als auch die Prüfung theoretisch ebenfalls plausibler entgegengesetzter Effekte fachbezogener Kompetenzen auf die spätere analytische Problemlösekompetenz ermöglichen, wie sie beispielsweise im internationalen Bericht zum Problemlösen in PISA 2012 angesprochen werden (OECD, 2014, Kap. 2). Zum anderen würde dieses Design, hinreichend große Item- und Personenstichproben vorausgesetzt, auch den Einsatz elaborierterer Analyseverfahren wie Wachstumskurvenmodelle oder auch Latent-Change-Modelle erlauben (z. B. McArdle & Grimm, 2010; Schmiedek & Wolff, 2010). Derartige Untersuchungen sollten jedoch durch kleinere (quasi)experimentelle Studien begleitet werden, denn tragfähige Kausalinterpretationen sind letztendlich nur durch eine kontrollierte Manipulation der entsprechenden Variablen im Rahmen experimenteller Designs möglich (Maxwell & Delaney, 2004).

Darüber hinaus sollte geprüft werden, ob die Berücksichtigung unterschiedlicher Arten von Problemstellungen ein differenzierteres Bild der Bedeutsamkeit der analytischen Problemlösekompetenz für zukünftige fachbezogene Kompetenzen liefern kann. Die Ergebnisse von Studie 2 hinsichtlich differenzieller Zusammenhänge für Problemstellungen der Art *Systeme analysieren und entwerfen* und *Fehler suchen* mit fachbezogenen Kompetenzen im Querschnitt lassen dies vermuten. Die Berücksichtigung dieses Aspekts der faktoriellen Validität der Modellierung der analytischen Problemlösekompetenz war im Rahmen von Studie 3 nicht möglich, da für diese Dimensionen keine entsprechenden Plausible Values vorlagen. Ferner wäre zu prüfen, inwiefern analytische Problemlösekompetenz auch ein Inkrement gegenüber fluiden Fähigkeiten bei der Vorhersage stärker curricular orientierter Kompetenztests aufweist (für die Mathematik z. B. VERA-8; siehe Leutner, Fleischer, Spoden & Wirth, 2007).<sup>10</sup> Dies betrifft insbesondere die naturwissenschaftlichen Kompetenzen in PISA, für die es in Deutschland – anders als bei der Mathematikkompetenz – keine schulfachspezifische Entsprechung gibt. Hier wäre interessant zu prüfen, inwiefern es gegebenenfalls differenzielle Effekte auf fachbezogene Kompetenzen in der Chemie, der Physik und der Biologie gibt. Dies könnte weitere Anhaltspunkte für die Art und Weise liefern,

---

<sup>10</sup> Für weitere Erläuterungen sei auf die entsprechende Webseite des IQB verwiesen: <https://www.iqb.hu-berlin.de/vera>.

wie das kognitive Potenzial, welches sich – gemäß der Potenzialausschöpfungshypothese – beim analytischen Problemlösen manifestiert, zum Aufbau fachbezogener Kompetenzen in Schule und Unterricht didaktisch genutzt werden kann.

## 4.7 Literatur

- Ackerman, P. L. & Beier, M. E. (2006). Determinants of domain knowledge and independent study learning in an adult sample. *Journal of Educational Psychology, 98*, 366–381. doi: 10.1037/0022-0663.98.2.366
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M. & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGraw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Dordrecht: Springer. doi: 10.1007/978-94-007-2324-5\_2
- Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (Hrsg.). (2006). *Bildungsstandards Mathematik: konkret*. Berlin: Cornelson.
- Buchwald, F., Fleischer, J. & Leutner, D. (2015). A field experimental study of analytical problem solving competence – Investigating effects of training and transfer. *Thinking Skills and Creativity, 18*, 18–31. doi: 10.1016/j.tsc.2015.04.009
- Buchwald, F., Fleischer, J., Rumann, S., Wirth, J. & Leutner, D. (2017). Training in components of problem-solving competence: An experimental study of aspects of the cognitive potential exploitation hypothesis. In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 315–331). Berlin: Springer. doi: 10.1007/978-3-319-50030-0\_19
- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Carstensen, C. H. (2006). Technische Grundlagen für die Messwiederholung. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Untersuchung zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 309–323). Münster: Waxmann.
- Carstensen, C. H., Knoll, S., Rost, J. & Prenzel, M. (2004). Technische Grundlagen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 371–387). Münster: Waxmann.



- Clemmer, S. C., Klifman, T. J. & Bradley-Johnson, S. (1992). Long-term predictive validity of the cognitive ability scales. *Journal of Psychoeducational Assessment*, 10, 265–275. doi: 10.1177/073428299201000306
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Cronbach, L. J. & Gleser, G. C. (1965). Psychological tests and personnel decisions. Urbana, IL: University of Illinois Press.
- Deary, I. J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21. doi: 10.1016/j.intell.2006.02.001
- Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. N., Marchione, K. & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, 43, 1460–1473. doi: 10.1037/0012-1649.43.6.1460
- Fleischer, J., Wirth, J., Rumann, S. & Leutner, D. (2010). Strukturen fächerübergreifender und fachlicher Problemlösekompetenz – Analyse von Aufgabenprofilen. *Zeitschrift für Pädagogik*, 56. Beiheft, 239–248.
- Fischer, A., Greiff, S., Wüstenberg, S., Fleischer, J., Buchwald, F. & Funke, J. (2015). Assessing analytic and interactive aspects of problem solving competency. *Learning and Individual Differences*, 39, 172–179. doi: 10.1016/j.lindif.2015.02.008
- Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J. & Csapó, B. (2013). Complex problem solving in educational settings — Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, 105, 364–379. doi: 10.1037/a0031856
- Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 35–72). Hillsdale, NJ: Erlbaum.
- Hany, E. A. (1997). Literaturüberblick über den Einfluß der vorschulischen Entwicklung auf die Entwicklung im Grundschulalter. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 391–404). Weinheim: Beltz.

- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Berlin: Springer. doi: 10.1007/978-3-642-20072-4\_7
- Heller, K. A. & Perleth, C. (2000). *KFT 4-12+R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Hogrefe.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistung. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule: Enzyklopädie der Psychologie, Serie Pädagogische Psychologie* (Bd. 3, S. 71–176). Göttingen: Hogrefe.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan & J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.
- Hunsley, J. & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological and statistical issues. *Psychological Assessment*, 15(4), 446–455. doi: 10.1037/1040-3590.15.4.446
- Kline, P. (1991). *Intelligence. The psychometric view*. London: Routledge.
- Köller, O. & Baumert, J. (2008). Entwicklung schulischer Leistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6. Aufl., S. 735–768). Weinheim: Beltz.
- Leeson, P., Ciarrochi, P. L. & Heaven, P. C. L. (2008). Cognitive ability, personality, and academic performance in adolescence. *Personality and Individual Differences*, 45, 630–635. doi: 10.1016/j.paid.2008.07.006
- Leutner, D., Fleischer, J., Spoden, C. & Wirth, J. (2007). Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik. *Zeitschrift für Erziehungswissenschaft, Sonderheft 8*, 149–167.
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147–175). Münster: Waxmann.
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2005). Die Problemlösekompetenz in den Ländern der Bundesrepublik Deutschland. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche* (S. 125–146)? Münster: Waxmann.

- MacKinnon, D. P. Lockwood, C. M., Hoffman, J. M., West, S. G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83–104. doi: 10.1037//1082-989X.7.1.83
- Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford, UK: University Press.
- Maxwell, S. E. & Delaney, H. D. (2004). *Designing experiments and analyzing data. A model comparison perspective*. Mahwah, NJ: LEA.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal Data. *Annual Review of Psychology*, 60, 577–605. doi: 10.1146/annurev.psych.60.110707.163612
- McArdle, J. J. & Grimm, K. J. (2010). Five steps in latent curve and latent change score modeling with longitudinal data. In K. van Montfort, J. H. L. Oud & A. Satorra (Eds.), *Longitudinal research with latent variables* (pp. 245–273). Heidelberg: Springer. doi: 10.1007/978-3-642-11760-2\_8
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. doi: 10.1037/0003-066X.50.9.741
- OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing.
- OECD. (2004). *Problem solving for tomorrow's world: First measurements of cross-curricular competencies from PISA 2003*. Paris: OECD Publishing.
- OECD. (2005). *PISA 2003: Technical report*. Paris: OECD Publishing.
- OECD. (2014). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems (Volume V)*. Paris: OECD Publishing.
- PISA-Konsortium Deutschland (Hrsg.). (2006a). *PISA 2003. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- PISA-Konsortium Deutschland (Hrsg.). (2006b). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.

- Prenzel, M., Carstensen, C. H., Schöps, K. & Maurischat, C. (2006). Die Anlage des Längsschnitts bei PISA 2003. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Untersuchung zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 29–62). Münster: Waxmann.
- Rohde, T. E. & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35, 83–92. doi: 10.1016/j.intell.2006.05.004
- Rost, D. H. (2009). *Intelligenz Fakten und Mythen*. Weinheim: Beltz.
- Rumann, S., Fleischer, J., Stawitz, H., Wirth, J. & Leutner, D. (2010). Vergleich von Profilen der Naturwissenschafts- und Problemlöse-Aufgaben der PISA 2003-Studie. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 315–327.
- Schmiedek, F & Wolff, J. K. (2010). Latente Wachstumskurvenmodelle. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 1017–1029). Wiesbaden: VS. doi: 10.1007/978-3-531-92038-2
- Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, 23, 153–158.
- Sonnleitner, P., Keller, U., Martin, R. & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, 41, 289–305. doi: 10.1016/j.intell.2013.05.002
- Stern, E. (2001a). Intelligence, prior knowledge, and learning. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 7670–7674). Amsterdam: Elsevier. doi: 10.1016/B0-08-043076-7/02383-4
- Stern, E. (2001b). Intelligenz, Wissen, Transfer und der Umgang mit Zeichensystemen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 163–204). Lengerich: Pabst.
- Stern, E. (2003). Lernen ist der mächtigste Mechanismus der kognitiven Entwicklung: Der Erwerb mathematischer Kompetenzen. In W. Schneider & M. Knopf (Hrsg.), *Entwicklung, Lehren und Lernen – Zum Gedenken an Franz Emanuel Weinert* (S. 207–217). Göttingen: Hogrefe.

Süß, H.-M. (2001). Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 109–135). Lengerich: Pabst.

Westat (2012). *WesVar 5* [Computer software]. Rockville, MD: Westat.

## **5 STUDIE 4: EFFEKTE DER KONTEXTUELLEN EINKLEIDUNG VON TESTAUFGABEN AUF DIE SCHÜLERLEISTUNGEN IM ANALYTISCHEN PROBLEMLÖSEN UND IN DER MATHEMATIK**

### **5.1 Einleitung**

In den ersten drei Studien dieser Arbeit standen Fragen der faktoriellen (Kapitel 2), diskriminanten (Kapitel 3), prognostischen und inkrementellen Validität (Kapitel 4) als Aspekte der Konstruktvalidität (Messick, 1995) der Modellierung analytischer Problemlösekompetenz im Zentrum. Dabei wurde auf verschiedene Datensätze der PISA-Studie 2003 zurückgegriffen. Im Kontext von PISA werden Kompetenzen allgemein als „prinzipiell erlernbare, mehr oder minder bereichsspezifische Kenntnisse, Fertigkeiten und Strategien“ definiert (Baumert, Stanat & Demmrich, 2001, S. 22). Diese Definition des Kompetenzbegriffs steht in wesentlichen Punkten in Übereinstimmung mit der Verwendung des Kompetenzbegriffs sowohl in anderen Schulleistungsstudien (z. B. TIMSS, IGLU/PIRLS) als auch in psychologischen und erziehungswissenschaftlichen Studien im Rahmen der Bildungsforschung (siehe Fleischer, Koeppen, Kenk, Klieme & Leutner, 2013; Leutner, Fleischer, Grünkorn & Klieme, 2017; Klieme, Hartig & Rauch, 2008). Einer Empfehlung von Weinert (2001) folgend wird damit auf kognitive Aspekte fokussiert, während motivationale und emotionale Faktoren ausgeschlossen werden. In der Regel erfolgt diese Einschränkung aus pragmatischen Erwägungen und nicht aus dem Grund, dass die prinzipielle Bedeutung motivationaler Einstellungen und Orientierungen sowie emotionaler Faktoren für kompetentes Handeln negiert wird (Fleischer et al., 2013). Studie 4 erweitert nun in diesem Sinne die Perspektive der Studien 1 bis 3, indem die Bedeutung motivationaler und emotionaler Faktoren – die sowohl Einfluss auf fächerübergreifende als auch fachbezogene Kompetenzen haben können (Mayer, 1998; Scherer & Gustafsson, 2015; Stankov, Lee, Luo & Hogan, 2012; Wang & Eccles, 2013) – für erfolgreiches analytisches Problemlösen in den Blick genommen wird.

Wie bereits in Abschnitt 1.3.3 dargestellt, zeigen die PISA 2003-Ergebnisse für Schülerinnen und Schüler in Deutschland eine deutliche Diskrepanz zwischen der (im internationalen Vergleich hohen) Kompetenz im analytischen Problemlösen und den (im internationalen Vergleich geringeren) fachbezogenen Kompetenzen, unter anderem in der Mathematik (OECD, 2004, Kap. 3). Diese Diskrepanz wird im internationalen und im nationalen Berichtsband zu PISA 2003 im Sinne einer *Potenzialausschöpfungshypothese* interpretiert, wonach Schülerinnen und Schüler in Deutschland über kognitives Potenzial verfügen, welches in der Schule nicht hinreichend zum *Aufbau* fachbezogener Kompetenzen genutzt wird (Leutner, Klieme, Meyer & Wirth, 2004; OECD, 2004, Kap. 3). Ebenso könnte jedoch auch von einer mangelnden *Nutzung* tatsächlich erworbener und bereits vorhandener fachbezogener Kompetenzen durch die Schülerinnen und Schüler während der *Testbearbeitung* (Nutzungsdefizit) ausgegangen werden. Dieser Hypothese, im Folgenden als *Potenzialnutzungshypothese* bezeichnet, wird in Studie 4 anhand von zwei experimentellen Untersuchungen nachgegangen.

Ziel von Studie 4 ist es zum einen zu prüfen, ob von einem Nutzungsdefizit bei der Bearbeitung von PISA-Testaufgaben ausgegangen werden kann. Zum anderen wird untersucht, welche Faktoren dafür verantwortlich sein könnten, dass Schülerinnen und Schüler ihr kognitives Potenzial und ihre Kompetenzen bei der Bearbeitung von Testaufgaben nicht erfolgreich nutzen. Die Studien zielen damit auf mögliche Erklärungsansätze für die bei PISA 2003 gefundene Diskrepanz zwischen analytischer Problemlösekompetenz und fachbezogenen Kompetenzen, unter anderem in der Mathematik.

## 5.2 Theoretischer Hintergrund

### 5.2.1 Analytische Problemlösekompetenz und mathematische Kompetenz in PISA 2003

Problemlösen wird zum einen als *fächerübergreifende* Kompetenz betrachtet, bei der die zur Lösung notwendigen Ressourcen nicht einer einzelnen Wissensdomäne zugeordnet werden können und die auch für den beruflichen Erfolg eine wichtige Rolle spielt (OECD, 2004; siehe auch Binkley et al., 2012; Greiff, Kretschmar & Leutner, 2014). Neben dieser fächerübergreifenden Perspektive wird Problemlösen auch als *fachspezifische* Kompetenz

betrachtet, zum Beispiel in der Mathematik (z. B. Reiss & Törner, 2007; Schoenfeld, 1985). Fächerübergreifende Problemlösekompetenz wurde in der PISA-Studie 2003 in Form des analytischen Problemlösens und mathematisches Problemlösen in Form der *Mathematical Literacy* erfasst. Der PISA-Test zum analytischen Problemlösen zielt darauf ab, zu erfassen, „wie gut Schülerinnen und Schüler [...] darauf vorbereitet sind, anwendungsbezogene fächerübergreifende Problemstellungen zu erkennen, zu verstehen und zu lösen“ (Leutner et al., 2004, S. 147). Mathematical Literacy als Schlüsselbegriff der Konzeption des PISA-Mathematiktests bezieht sich auf die Frage, „inwieweit Schülerinnen und Schüler [...] Mathematik in problemhaltigen außer- oder innermathematischen Situationen flexibel anwenden können“ (Blum et al., 2004, S. 47).

Eine Beispielaufgabe aus dem PISA-Problemlösetest, wie sie auch in der vorliegenden Untersuchung verwendet wurde, stellt die Aufgabe „Anschlusszüge“ dar, bei der es darum geht, die schnellste und kostengünstigste Verbindung zwischen zwei U-Bahn-Stationen zu bestimmen, wobei es notwendig ist, von einer Linie in eine andere umzusteigen (siehe Abschnitt 1.3.3). Hierzu müssen Informationen aus der Abbildung eines U-Bahn-Liniennetzes sowie Informationen aus dem Aufgabentext entnommen und integriert werden (Leutner et al., 2004; für weitere Beispielaufgaben siehe OECD, 2004).

Eine Beispielaufgabe für den PISA-Mathematiktest stellt die ebenfalls in der vorliegenden Untersuchung verwendete Aufgabe „Exporte“ dar. Bei dieser Aufgabe muss in einem ersten Schritt der Gesamtwert der Exporte eines fiktiven Landes aus einer von zwei Grafiken abgelesen werden, wobei nur eine der Grafiken relevante Informationen enthält. In einem zweiten Schritt ist der Prozentwert einer Exportsumme näherungsweise zu berechnen, wobei Informationen aus beiden Grafiken verwendet werden müssen (für weitere Beispielaufgaben siehe Blum et al., 2004; OECD, 2003, Kap. 1).

Beide Kompetenzen weisen konzeptuelle Ähnlichkeiten auf, sowohl bezüglich der zur erfolgreichen Bearbeitung von analytischen Problemlöseaufgaben und Mathematikaufgaben erforderlichen Prozessschritte als auch bezüglich der im Ablauf des Lösungsprozesses relevanten Komponenten (Fleischer, Wirth, Rumann & Leutner, 2010), wie in Abschnitt 1.3 ausführlich dargestellt wurde. Diese konzeptuellen Ähnlichkeiten zwischen analytischer



Problemlösekompetenz und Mathematikkompetenz schlagen sich empirisch in vergleichsweise hohen (messfehlerbereinigten) Korrelationen zwischen beiden Kompetenzbereichen in PISA 2003 nieder (je nach zugrunde gelegter Stichprobe:  $r = .89$ , OECD, 2005, Kap. 13; bzw.  $r = .90$ , Leutner et al., 2004).

Durch die Skalierung der analytischen Problemlöseskala und der Mathematikskala in PISA 2003, jeweils auf einer Metrik von  $M = 500$  und  $SD = 100$  über alle OECD-Teilnehmerstaaten hinweg (internationale PISA-Metrik), lassen sich beide Kompetenzen relativ zum OECD-Durchschnitt miteinander vergleichen (Leutner et al., 2004; OECD, 2004, Kap. 4). Hierbei zeigt sich in Deutschland ein beachtlicher Niveauunterschied: Die (im internationalen Vergleich hohe) analytische Problemlösekompetenz der Schülerinnen und Schüler übersteigt die (im internationalen Vergleich durchschnittliche) mathematische Kompetenz um 10 Punkte auf der internationalen PISA-Metrik. Lediglich in zwei der insgesamt 29 getesteten Staaten fällt diese Differenz zugunsten der analytischen Problemlösekompetenz noch größer aus (Japan: 13 Punkte; Ungarn: 11 Punkte). Dieses Ergebnis wird im internationalen und im nationalen Berichtsband zu PISA 2003 (Leutner et al., 2004; OECD, 2004, Kap. 4) als mangelnde Ausschöpfung des beim Problemlösen offensichtlich vorhandenen kognitiven Potenzials zum Aufbau fachbezogener Kompetenzen interpretiert (Potenzialausschöpfungshypothese). Die entgegengesetzte Differenz zugunsten der Mathematikkompetenz, wie sie beispielsweise in den Niederlanden oder Island festzustellen ist, wird hingegen als potenzieller Indikator für erfolgreichen Mathematikunterricht interpretiert (OECD, 2004, Kap. 4).

### **5.2.2 Kontext und kognitive Leistungen – Moderierende Effekte des akademischen Selbstkonzepts und der Leistungsangst**

Die zu untersuchende Potenzialnutzungshypothese geht, wie eingangs beschrieben, im Gegensatz zur Potenzialausschöpfungshypothese davon aus, dass Schülerinnen und Schüler in einer konkreten, kontextualisierten Testsituation ihr kognitives Potenzial und ihre Fachkompetenz nicht hinreichend zur Bearbeitung der gestellten Testaufgaben nutzen. Zahlreiche Studien haben gezeigt, dass der Kontext, in dem eine kognitive Leistung zu erbringen ist, einen Einfluss auf die Qualität dieser Leistung haben kann (z. B. Barab &

Plucker, 2002; Blessing & Ross, 1996; Kirsh, 2009; Kotovsky, Hayes & Simon, 1985). In der Forschung lassen sich hierzu grob zwei Forschungsstränge unterscheiden. Zum einen werden kontextfreie Aufgaben (z. B. Arithmetikaufgaben) mit strukturgleichen Aufgaben verglichen, die in einem bestimmten Kontext eingebunden sind (z. B. mathematische Textaufgaben). Hierbei zeigt sich, dass kontextualisierte Aufgabenstellungen häufig höhere Lösungsquoten aufweisen als kontextfreie (z. B. Clements, 1980; van den Heuvel-Panhuizen, 2005). Zum anderen werden Aufgaben oder Problemstellungen in einem Anwendungskontext mit strukturgleichen Aufgaben oder Problemstellungen in einem anderen Anwendungskontext verglichen. Hierbei zeigt sich, dass kognitive Operationen, welche in einem bestimmten Kontext beherrscht werden, häufig nicht oder nur unter bestimmten Bedingungen in andere Kontexte transferiert werden (Barnett & Ceci, 2002; Lave, 1988). Darüber hinaus können unterschiedliche Kontexte unterschiedliche Lösungsstrategien nahelegen (Bjorklund & Rosenblum, 2002; Zöttl, Heinze & Reiss, 2007). So konnten in der klassischen Untersuchung von Carraher, Carraher und Schliemann (1985) zur Rechenfähigkeit von brasilianischen Straßenkindern die Kinder fast alle ihnen gestellten Aufgaben im Kontext des Straßenhandels erfolgreich lösen (98 %). Wurden dieselben Aufgaben jedoch als Textaufgaben präsentiert, sank die Lösungsquote auf 74 Prozent, und wurden die Aufgaben vollständig aus ihrem Kontext gelöst und als reine Mathematikaufgaben präsentiert, sank die Lösungsquote weiter auf 37 Prozent.<sup>11</sup>

Nicht allein die zur Lösung einer Aufgabe notwendigen Kompetenzen entscheiden also darüber, ob eine Aufgabe korrekt gelöst wird. Auch die Kontexte, in welche die Aufgaben eingebunden sind, können Einfluss auf die Lösungshäufigkeit dieser Aufgaben haben. Allerdings lässt sich nicht eindeutig sagen, welche Art von Kontext per se leistungsförderlich ist. Nach van den Heuvel-Panhuizen (2005) sollte der Aufgabenkontext zugänglich, einladend und bedeutsam für die aufgabenbearbeitende Person sein. Diese Nähe des Aufga-

---

<sup>11</sup> Der Vollständigkeit halber sei in diesem Zusammenhang auch auf Arbeiten zum Vergleich unterschiedlicher Repräsentationsformate im Rahmen diagnostischer Entscheidungen hingewiesen. So zeigen beispielsweise Gigerenzer und Hoffrage (1995) einen Vorteil der Präsentation von bedingten Wahrscheinlichkeiten in Form von natürlichen Häufigkeiten, die auf dem Zählen tatsächlich beobachtbarer Ereignisse basieren (einem eher realitätsnahen Kontext), gegenüber der Darstellung in Form von Prozentangaben (einem eher mathematischen Kontext).

benkontextes zur Lebenswirklichkeit von Schülerinnen und Schülern sollte bei den analytischen Problemlöseaufgaben aus PISA in stärkerem Maße gegeben sein als bei den Mathematikaufgaben (vgl. Fleischer et al., 2010).

Bei dem durch die Potenzialnutzungshypothese postulierten Kontexteffekt stellt sich die Frage, welche Faktoren dafür verantwortlich sind, dass Schülerinnen und Schüler ihr kognitives Potenzial nicht adäquat nutzen. Die Vermutung liegt nahe, dass unterschiedliche Kontexte wie ein alltäglicher Problemlösekontext oder ein fachbezogener mathematischer Kontext, in denen kognitive Leistungen zu erbringen sind, nicht nur unterschiedliche Lösungsstrategien nahelegen, sondern auch über die Aktivierung von kognitiv-motivationalen und emotionalen Faktoren differenziell auf kognitive Leistungen wirken. So sollte beispielsweise ein fachbezogener mathematischer Kontext auf Personen mit positiven kognitiv-motivationalen und emotionalen Dispositionen bezüglich der Mathematik eine andere Wirkung haben als auf Personen mit diesbezüglich negativen Dispositionen.

Die theoretische und praktische Bedeutung insbesondere des akademischen Selbstkonzepts und der Leistungsangst, die häufig am Beispiel von Schülerleistungen in einem mathematischen Kontext untersucht wird, konnte vielfach gezeigt werden (z. B. Ma, 1999; Möller & Köller, 2004). Wie im Folgenden dargestellt, können durch beide Variablen leistungsrelevante Indikatoren erklärt und vorhergesagt werden.

Domänenspezifische *akademische Fähigkeitsselbstkonzepte* stellen kognitive Repräsentationen der eigenen Fähigkeiten in einer bestimmten Domäne dar, zum Beispiel „Ich bin gut in Mathematik“ (vgl. Meyer, 1984). Zusammenhänge zwischen domänenspezifischen akademischen Selbstkonzepten und Indikatoren für Schulleistungen in derselben Domäne sind vielfach belegt. Die Korrelationen liegen in der Regel in einem Bereich zwischen  $r = .45$  und  $r = .70$  (Marsh, 1992; vgl. Valentine, DuBois, & Cooper, 2004). Akademische Selbstkonzepte sind zum einen das Ergebnis bisheriger schulischer Leistungen, zum anderen können akademische Selbstkonzepte – zum Beispiel über Interesse und Anstrengungsbereitschaft vermittelt – auch Einfluss auf zukünftige schulische Leistungen nehmen (z. B. Helmke & van Aken, 1995; Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005). So lässt sich vermuten, dass sich Personen mit einem geringen Selbstkonzept weniger anspruchsvolle

Teilziele setzen oder die Lösung eines Problems erst gar nicht in Angriff nehmen (vgl. Bouffard-Bouchard, Parent & Larivée, 1991).

*Leistungsangst* lässt sich definieren als „die Besorgtheit und Aufgeregtheit angesichts von Leistungsanforderungen, die als selbstwertbedrohlich eingeschätzt werden“ (Schwarzer, 2000, S. 105). Zusammenhänge von Leistungsangst mit Schulleistungen sind vielfach belegt (z. B. Chappell et al., 2005). So berichtet Ma (1999) in seiner Metaanalyse eine mittlere Korrelation von  $r = -.27$  für Mathematik. Ähnlich wie beim Zusammenhang von akademischem Selbstkonzept und Schulleistungen kann auch beim Zusammenhang von Leistungsangst und Schulleistungen von einer Wechselwirkung ausgegangen werden: Leistungsangst ist zum einen die Folge bisheriger Leistungen, zum anderen kann Leistungsangst jedoch auch Einfluss auf zukünftige Leistungen haben, beispielsweise vermittelt über aufgabenirrelevante Kognitionen und reduzierte Anstrengungsbereitschaft (Cassady & Johnson, 2002; Schwarzer, 2000).

Für das akademische Selbstkonzept und Leistungsangst in derselben Domäne zeigen sich in der Regel hohe Korrelationen (z. B.  $r = -.65$ ; Ferla, Valcke & Cai, 2009). Dennoch werden in der Forschung häufig beide Konstrukte simultan als Prädiktoren schulischer Leistungen untersucht (z. B. Pajares & Graham, 1999).

### 5.3 Fragestellungen und Hypothesen

Ziel von Studie 4 ist es zum einen zu überprüfen, ob von einem Nutzungsdefizit bei Schülerinnen und Schülern in Deutschland bei der Bearbeitung von PISA-Testaufgaben ausgegangen werden kann. Wie im vorherigen Abschnitt dargestellt, kann die kontextuelle Einkleidung von Testaufgaben einen Einfluss auf die Lösungshäufigkeit dieser Aufgaben haben. Es soll daher zunächst der Frage nachgegangen werden, ob unterschiedliche kontextuelle Einkleidungen von ansonsten identischen PISA-Testaufgaben einen Einfluss auf die Schülerleistungen haben können.

Zum anderen verfolgt die Untersuchung das Ziel, Faktoren zu identifizieren, die Schülerinnen und Schüler daran hindern, ihr kognitives Potenzial bei der Bearbeitung von Test-

aufgaben erfolgreich zu nutzen. Wie zuvor dargestellt, können das akademische Selbstkonzept und die Leistungsangst Einfluss auf Schülerleistungen nehmen. Testaufgaben, die in einem mathematischen Kontext präsentiert werden und von Lernenden daher als Mathematikaufgaben wahrgenommen werden, und Aufgaben, die in einem alltäglichen Problemlösekontext präsentiert werden und von Lernenden daher als Problemlöseaufgaben wahrgenommen werden, sollten demnach je nach Ausprägung des mathematischen Selbstkonzepts und der Mathematikangst unterschiedliche Lösungsquoten aufweisen. Es wird somit der Frage nachgegangen, ob unterschiedliche kontextuelle Einkleidungen von PISA-Testaufgaben in Abhängigkeit vom mathematischen Selbstkonzept und von der Mathematikangst differenziell auf die Schülerleistungen wirken.<sup>12</sup> Zur Untersuchung dieser Fragestellungen wurden zwei experimentelle Studien durchgeführt.

In Studie 4a wurden PISA-Mathematikaufgaben in einem *between-subjects*-Design in zwei unterschiedlichen Kontextbedingungen (fachbezogener mathematischer Kontext versus fächerübergreifender Problemlösekontext) präsentiert. Es wird hierbei von einem Haupteffekt der kontextuellen Einkleidung auf die Schülerleistungen ausgegangen. Gemäß der Potenzialnutzungshypothese wird erwartet, dass die Schülerleistungen in den Mathematikaufgaben unter der Mathematikbedingung geringer ausfallen als unter der Problemlösebedingung (Hypothese 1a). Darüber hinaus wird von Moderationseffekten des mathematischen Selbstkonzepts und der Mathematikangst ausgegangen, das heißt, es wird erwartet, dass der Effekt der kontextuellen Einkleidung auf die Schülerleistungen in Mathematik besonders ausgeprägt ist bei Schülerinnen und Schülern mit geringem mathematischem Selbstkonzept (Hypothese 2a) sowie bei Schülerinnen und Schülern mit hoher Mathematikangst (Hypothese 3a).

In Studie 4b wurde geprüft, ob analoge Effekte auch für analytische Problemlöseaufgaben gezeigt werden können. Hierzu wurden PISA-Problemlöseaufgaben in einem *between-subjects*-Design entsprechend dem Vorgehen in Studie 4a in zwei unterschiedlichen Kontextbedingungen (fachbezogener mathematischer Kontext versus fächerüber-

---

<sup>12</sup> Es wurde zudem geprüft, ob unterschiedliche kontextuelle Einkleidungen ebenfalls in Abhängigkeit von Geschlecht und Migrationshintergrund differenziell auf die Schülerleistungen wirken. Da dies nicht der Fall ist, wird auf die Darstellung dieser Analysen verzichtet.

greifender Problemlösekontext) präsentiert. Auch hier wird erwartet, dass die Schülerleistungen im analytischen Problemlösen unter der Mathematikbedingung geringer ausfallen als unter der Problemlösebedingung (Hypothese 1b). Darüber hinaus wird wie in Studie 4a davon ausgegangen, dass der Effekt der kontextuellen Einkleidung auf die Schülerleistungen im analytischen Problemlösen bei Schülerinnen und Schülern mit geringem mathematischem Selbstkonzept (Hypothese 2b) sowie bei Schülerinnen und Schülern mit hoher Mathematikangst (Hypothese 3b) besonders ausgeprägt ist.

## 5.4 Methode

### 5.4.1 Stichproben

An Studie 4a nahmen  $N = 256$  Schülerinnen und Schüler (49.2 % weiblich) der 9. Jahrgangsstufe aus Nordrhein-Westfalen teil. Die Schülerinnen und Schüler rekrutierten sich aus 24 Klassen aus 9 Schulen der Schulformen Hauptschule, Realschule und Gymnasium. 192 Schülerinnen und Schüler (75.6 %) gaben als Familiensprache Deutsch an. An Studie 4b nahmen  $N = 259$  Schülerinnen und Schüler (45.6 % weiblich) ebenfalls der 9. Jahrgangsstufe aus Nordrhein-Westfalen teil. 187 Schülerinnen und Schüler (72.2 %) gaben als Familiensprache Deutsch an. Die Stichprobe rekrutierte sich aus denselben 24 Klassen der 9 Schulen der Studie 4a, wobei jede Schülerin und jeder Schüler nur an einer der beiden Studien teilnahm.

### 5.4.2 Instrumente

Zur Erfassung der Mathematikleistung in Studie 4a und der analytischen Problemlöseleistung in Studie 4b wurden jeweils fünf Testitems der entsprechenden PISA-Skalen der Erhebung 2003 ausgewählt (siehe die in Abschnitt 1.3.3 beschriebenen Beispielaufgaben). Für den Mathematiktest ( $M = 2.91$ ,  $SD = 2.08$ ) ergab sich eine interne Konsistenz von Cronbachs  $\alpha = .65$  und für den analytischen Problemlösetest ( $M = 2.86$ ,  $SD = 1.50$ ) ein Wert von Cronbachs  $\alpha = .50$ . Die relativ niedrigen internen Konsistenzen der beiden Skalen dürften auf die geringe Anzahl an Items pro Skala zurückzuführen und kein Resultat etwaiger

Verzerrungen der verwendeten Stichproben sein. Dafür spricht, dass eine eigene Reskalierung der verwendeten Items anhand des deutschen PISA 2003-Datensatzes nahezu identische Ergebnisse lieferte.

Die beiden Kontextbedingungen „fachbezogener mathematischer Kontext“ und „fächerübergreifender Problemlösekontext“ wurden in beiden Studien jeweils durch unterschiedliche Testhefte umgesetzt. Im Falle der mathematischen Kontextbedingung bestanden die Testhefte aus einem Cover, welches auf Inhalte der Mathematik verwies, gefolgt von einem einführenden Text, der die allgemeine Wichtigkeit und die Bedeutung der Mathematikkompetenz für den späteren beruflichen Erfolg thematisierte. Im Anschluss folgte, als einführende Aufgabe, ein Item aus dem PISA-2003-Mathematiktest. Im Falle des Problemlösekontexts bestanden die Testhefte aus einem Cover, welches auf das Lösen alltäglicher Probleme verwies, gefolgt von einem einführenden Text, der die Wichtigkeit und die Bedeutung der allgemeinen Problemlösekompetenz für den späteren beruflichen Erfolg thematisierte. Daran schloss sich, als einführende Aufgabe, ein Item aus dem PISA-2003-Problemlösetest an. Das Cover, der einführende Text sowie die einführende Aufgabe stellten gemeinsam jeweils die kontextuelle Einbettung „fachbezogener mathematischer Kontext“ versus „fächerübergreifender Problemlösekontext“ dar, in welcher der Mathematiktest in Studie 4a und der analytische Problemlösetest in Studie 4b präsentiert wurden (siehe Anhang A und B). Sowohl der Mathematiktest als auch der analytische Problemlösetest unterschieden sich in beiden kontextuellen Einbettungen nicht voneinander, die Testaufgaben waren jeweils identisch.

Die Erfassung des mathematischen Selbstkonzepts erfolgte mithilfe der entsprechenden Skala des *DISK-Gitters* (Rost & Sparfeldt, 2002), bestehend aus acht Items (z. B. „Ich weiß in Mathematik die Antwort auf eine Frage schneller als die anderen“) mit jeweils 6-stufiger Likert-Antwortskala (1 = *trifft gar nicht zu* bis 6 = *trifft genau zu*;  $M = 3.46$ ,  $SD = 1.22$ ; Cronbachs  $\alpha = .92$ ). Zur Erfassung der Mathematikangst wurde die Skala *Ängstlichkeit in Mathematik* des Schülerfragebogens aus PISA 2003 (PISA-Konsortium Deutschland, 2006, Teil II, Kap. 5) bestehend aus fünf Items (z. B. „Beim Lösen von Mathematikaufgaben fühle ich mich hilflos“) mit jeweils 4-stufiger Likert-Antwortskala eingesetzt (1 = *stimmt überhaupt nicht* bis 4 = *stimmt ganz genau*;  $M = 2.08$ ,  $SD = 0.71$ ; Cronbachs

$\alpha = .85$ ). Darüber hinaus wurden die Mathematiknote im letzten Zeugnis sowie die Schulform als Kontrollvariablen erhoben.

### 5.4.3 Untersuchungsdesign und Durchführung

In Studie 4a wurde der Mathematiktest in einem *between-subjects*-Design entweder in der Kontextbedingung Mathematik oder in der Kontextbedingung Problemlösen zur Bearbeitung vorgelegt. Die Testhefte bestanden aus einem Cover, dem einführenden Text sowie der einführenden Aufgabe, die gemeinsam die kontextuelle Einbettung darstellten. Im Anschluss war der Mathematiktest zu bearbeiten. Hierfür waren 20 Minuten vorgesehen. Daran anschließend erfolgten die Beantwortung der Skalen zum mathematischen Selbstkonzept und zur Mathematikangst sowie die Erhebung der Mathematiknote. Die Testhefte wurden zufällig auf die Schülerinnen und Schüler jeder Klasse verteilt, und die gesamte Erhebung wurde innerhalb der Dauer einer regulären 45-minütigen Schulstunde durchgeführt. In Studie 4b wurde der analytische Problemlösetest in einem *between-subjects*-Design entweder in der Kontextbedingung Mathematik oder in der Kontextbedingung Problemlösen zur Bearbeitung vorgelegt. Aufbau und Durchführung der Studie 4b erfolgten analog zu Studie 4a.

### 5.4.4 Statistische Analysen

Da die Stichproben beider Studien eine geclusterte Struktur aufweisen (Schülerinnen und Schüler geclustert in Schulklassen), wurde zunächst die Intraklassenkorrelation (ICC) der Testleistungen in beiden Studien bestimmt. Die ICC gibt den Anteil der Gesamtvarianz der Leistungen an, der allein auf die Gruppenzugehörigkeit (Schulklasse) der Schülerinnen und Schüler zurückgeführt werden kann (vgl. Goldstein, 2011). Für die Mathematikleistung in Studie 4a ergab sich eine ICC von  $\rho = .30$  und für die analytische Problemlöseleistung in Studie 4b eine ICC von  $\rho = .23$ , wodurch eine mehrebenenanalytische Auswertung der Daten angebracht erschien, da bereits wesentlich geringere ICCs zu einer Inflation des Alpha-Fehlers bei der Signifikanztestung führen können (vgl. Goldstein, 2011). Die Analyse der Daten erfolgte mit dem Programm HLM 7 (Raudenbush, Bryk, Cheong, Congdon & du Toit,



2011), wodurch korrekte Schätzungen der Standardfehler der Regressionskoeffizienten und damit eine korrekte Signifikanzprüfung gewährleistet sind. Alle Modelle wurden als *random-intercept*-Modelle geschätzt, bei denen neben den Prädiktoren auf Ebene 1 (Schülerebene) die Schulform auf Ebene 2 (Klassenebene) kontrolliert wurde. Die abhängigen Variablen Mathematikleistung (Studie 4a) und analytische Problemlöseleistung (Studie 4b) wurden in den Analysen z-standardisiert, ebenso die Prädiktoren mathematisches Selbstkonzept und Mathematikangst. Die Variable Kontext wurde dummy-kodiert.

## 5.5 Ergebnisse

### 5.5.1 Ergebnisse Studie 4a

Die Schülerinnen und Schüler in beiden Kontextgruppen unterschieden sich weder hinsichtlich ihrer Mathematiknote ( $t_{(254)} = 0.54$ ;  $p = .59$ ; Cohens  $d = 0.07$ ) noch bezüglich ihres mathematischen Selbstkonzepts ( $t_{(254)} = -0.93$ ;  $p = .36$ ; Cohens  $d = 0.12$ ) und ihrer Mathematikangst ( $t_{(253)} = -0.11$ ;  $p = .91$ ; Cohens  $d = 0.01$ ), was für eine erfolgreiche zufällige Gruppenzuordnung spricht. Tabelle 5.1 stellt die Mittelwerte und Standardabweichungen der in den folgenden Analysen betrachteten Variablen dar. Die Korrelation von mathematischem Selbstkonzept und Mathematikangst bewegte sich mit  $r = -.69$  in der zu erwartenden Größenordnung.

**Tabelle 5.1: Mittelwerte und Standardabweichungen (in Klammern) der in den Mehrebenenanalysen in Studie 4a verwendeten Variablen**

Variablen	Kontext: Mathematik	Kontext: Problemlösen
Mathematikleistung	2.68 (2.17)	3.15 (1.97)
Mathematisches Selbstkonzept	3.41 (1.21)	3.27 (1.27)
Mathematikangst	2.13 (0.74)	2.12 (0.74)

Anmerkungen.  $N = 256$ .

Tabelle 5.2 gibt die Regressionskoeffizienten der Mehrebenenanalysen zur Prüfung der Hypothesen 1a bis 3a wieder. In allen Modellen wurde die Schulformzugehörigkeit auf Ebene 2 (in Tabelle 5.2 nicht dargestellt) als Kontrollvariable verwendet.

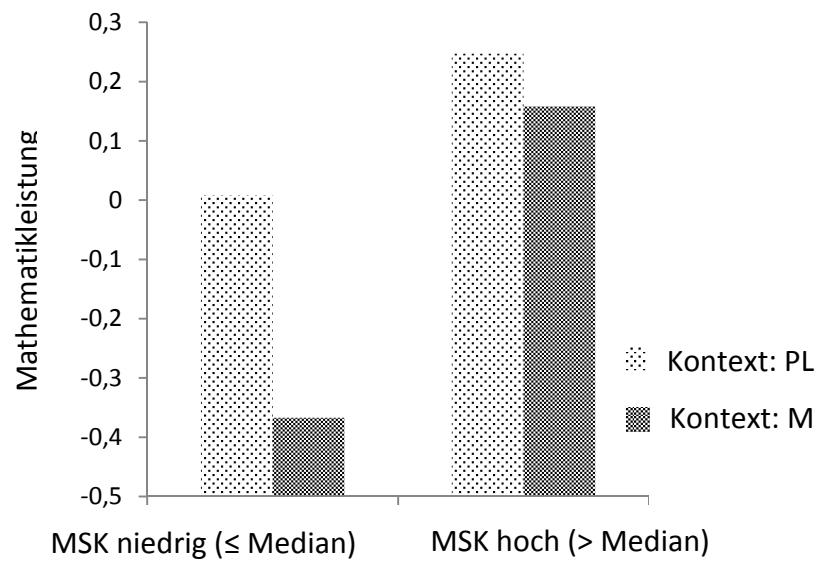
In Modell 1a zeigte sich ein signifikanter negativer Effekt der kontextuellen Einbettung der Mathematikaufgaben auf die Leistungen der Schülerinnen und Schüler in diesen Aufgaben ( $b = -.25$ ). Hypothese 1a konnte damit bestätigt werden. Zur Prüfung von Hypothese 2a wurden die Variable Mathematisches Selbstkonzept (MSK) sowie ein Term für die Interaktion von MSK und der kontextuellen Einkleidung (MSK\*Kontext) in Modell 2a eingeführt (vgl. Aiken & West, 1991). Unter Kontrolle der beiden Haupteffekte zeigte sich ein signifikanter Interaktionseffekt MSK\*Kontext ( $b = .21$ ). Getrennte Analysen für Schülerinnen und Schüler mit hohem versus niedrigem mathematischem Selbstkonzept (Mediansplit) verdeutlichen die Art dieser Interaktion. Während sich bei Schülerinnen und Schülern mit niedrigem mathematischen Selbstkonzept ein signifikanter negativer Effekt der mathematischen Kontextbedingung zeigte ( $b = -.42$ ,  $KI_{95} = -.71/-.13$ ), wurde der entsprechende Effekt bei Schülerinnen und Schülern mit hohem mathematischen Selbstkonzept nicht signifikant ( $b = -.22$ ,  $KI_{95} = -.50/.07$ ). Abbildung 5.1 (deskriptive Darstellung der Rohwerte) veranschaulicht diese Interaktion zudem grafisch. Für eine differenziertere Betrachtung wurde die Stichprobe anhand des mathematischen Selbstkonzepts zusätzlich in Quartile aufgeteilt. Dabei zeigte sich für Schülerinnen und Schüler mit sehr geringem mathematischen Selbstkonzept (1. Quartil) ein signifikanter Effekt ( $b = -0.63$ ,  $KI_{95}: -1.04/-0.22$ ), während sich für Schülerinnen und Schüler in den anderen drei Quartilen keine signifikanten Effekte zeigen ließen. Hypothese 2a konnte damit bestätigt werden. Zur Prüfung von Hypothese 3a wurden die Variable Mathematikangst (MA) sowie ein Term für die Interaktion von MA und kontextueller Einkleidung (MA\*Kontext) in Modell 3a eingeführt. Unter Kontrolle der jeweiligen Haupteffekte zeigte sich ein signifikanter Interaktionseffekt MA\*Kontext ( $b = -.17$ ). Getrennte Analysen für Schülerinnen und Schüler mit hoher versus niedriger Mathematikangst (Mediansplit) verdeutlichen auch hier die Art der Interaktion. Der negative Effekt der Kontextbedingung „Mathematik“ zeigte sich jedoch nur bei Schülerinnen und Schülern mit hoher Mathematikangst ( $b = -.53$ ,  $KI_{95}: -.83/-.23$ ), während sich bei Schülerinnen und Schülern mit niedriger Mathematikangst kein signifikanter Effekt zeigte ( $b = -.11$ ,  $KI_{95}: -.39/.17$ ). Abbildung 5.2 (deskriptive Darstellung der Rohwerte) veranschaulicht

diese Interaktion wiederum grafisch. Für eine differenziertere Betrachtung wurde die Stichprobe anhand der Mathematikangst zusätzlich in Quartile aufgeteilt. Dabei zeigte sich für Schülerinnen und Schüler mit sehr hoher Mathematikangst (4. Quartil) ein signifikanter Effekt ( $b = -0.92$ ;  $KI_{95} : -1.33/-0.51$ ), während sich für Schülerinnen und Schüler in den anderen drei Quartilen keine signifikanten Effekte zeigen ließen. Hypothese 3a konnte damit ebenfalls bestätigt werden.

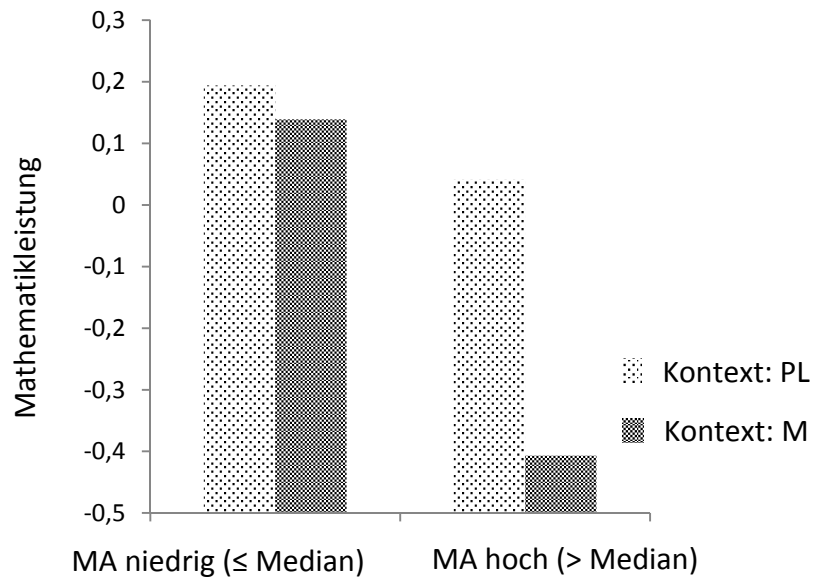
**Tabelle 5.2: Ergebnisse der Mehrebenenanalysen zur Vorhersage der Mathematikleistung (Studie 4a): Unstandardisierte Regressionskoeffizienten (95%-Konfidenzintervall in eckigen Klammern) und aufgeklärte Varianz**

	Modell 1a	Modell 2a	Modell 3a
Kontext (0=Problemlösen, 1=Mathematik)	-.25[-.45/-.04]	-.25[-.45/-.06]	-.23[-.42/-.03]
Mathematisches Selbstkonzept (MSK; z-standardisiert)		.15[.01/.29]	
Interaktion: MSK*Kontext		.21[.01/.41]	
Mathematikangst (MA; z-standardisiert)			-.17[-.30/-.03]
Interaktion: MA*Kontext			-.22[-.41/-.02]
$R^2$	.06	.09	.12

*Anmerkungen.*  $R^2$ : Anteil der aufgeklärten Varianz auf Individualebene; für alle Regressionskoeffizienten gilt:  $p < .05$ ;  $N = 256$ .



**Abbildung 5.1: Mittelwerte der Mathematikleistung (z-standardisiert) nach mathematischem Selbstkonzept (MSK; dichotomisiert am Median) und Kontextbedingung (Problemlösekontext [PL] versus Mathematikkontext [M]);  $N = 256$**



**Abbildung 5.2: Mittelwerte der Mathematikleistung (z-standardisiert) nach Mathematikangst (MA; dichotomisiert am Median) und Kontextbedingung (Problemlösekontext [PL] versus Mathematikkontext [M]);  $N = 256$**

### 5.5.2 Ergebnisse Studie 4b

Auch in Studie 4b unterschieden sich die Schülerinnen und Schüler in beiden Kontextgruppen weder hinsichtlich ihrer Mathematiknote ( $t_{(257)} = 0.02$ ;  $p = .98$  Cohens  $d < 0.01$ ) noch bezüglich ihres mathematischen Selbstkonzepts ( $t_{(257)} = -0.69$ ;  $p = .49$ ; Cohens  $d = 0.09$ ) und ihrer Mathematikangst ( $t_{(243)} = -1.54$ ;  $p = .13$ ; Cohens  $d = 0.19$ ), was für eine erfolgreiche zufällige Gruppenzuordnung spricht. Tabelle 5.3 stellt die Mittelwerte und Standardabweichungen der in den folgenden Analysen betrachteten Variablen dar. Die Korrelation von mathematischem Selbstkonzept und Mathematikangst war mit  $r = -.60$  etwas niedriger als in Studie 4a.

**Tabelle 5.3: Mittelwerte und Standardabweichungen (in Klammern) der in den Mehrebenenanalysen in Studie 4b verwendeten Variablen**

Variablen	Kontext: Mathematik	Kontext: Problemlösen
Problemlöseleistung	2.67 (1.61)	3.04 (1.38)
Mathematisches Selbstkonzept	3.63 (1.20)	3.53 (1.16)
Mathematikangst	2.10 (0.73)	1.97 (0.60)

Anmerkungen.  $N = 259$ .

Tabelle 5.4 stellt die Regressionskoeffizienten der Mehrebenenanalysen zur Prüfung der Hypothesen 1b bis 3b dar. In allen Modellen wurde die Schulformzugehörigkeit auf Ebene 2 (in Tabelle 5.4 nicht dargestellt) als Kontrollvariable verwendet.

In Modell 1b zeigte sich ein signifikanter negativer Effekt der kontextuellen Einbettung auf die analytischen Problemlöseleistungen ( $b = -.24$ ). Hypothese 1b konnte damit bestätigt werden. In Modell 2b zeigte sich für die Interaktion von mathematischem Selbstkonzept und Kontext (MSK\*Kontext) unter Kontrolle beider Haupteffekte ein signifikanter Effekt ( $b = .27$ ). Getrennte Analysen für Schülerinnen und Schüler mit hohem versus niedrigem mathematischem Selbstkonzept (Mediansplit) zeigten einen signifikanten negativen

Effekt der mathematischen Kontextbedingung bei Schülerinnen und Schülern mit geringem mathematischen Selbstkonzept ( $b = -.46$ ,  $KI_{95}: -.79/-13$ ). Für Schülerinnen und Schüler mit hohem mathematischen Selbstkonzept zeigte sich hingegen kein signifikanter Effekt ( $b = -.03$ ,  $KI_{95}: -.32/.26$ ). Eine differenziertere Betrachtung anhand von Quartilen zeigte nur für Schülerinnen und Schüler mit sehr geringem mathematischen Selbstkonzept (1. Quartil) einen signifikanten Effekt ( $b = -1.01$ ;  $KI_{95}: -1.45/-0.57$ ), während sich für die anderen drei Quartile keine signifikanten Effekte zeigen ließen. Hypothese 2b konnte damit bestätigt werden. In Modell 3b wurde auch die Interaktion von Mathematikangst und kontextueller Einkleidung (MA\*Kontext) unter Kontrolle beider Haupteffekte signifikant ( $b = -.42$ ). Getrennte Analysen für Schülerinnen und Schüler mit hoher versus niedriger Mathematikangst (Mediansplit) zeigten einen signifikanten negativen Effekt der Kontextbedingung „Mathematik“ bei Schülerinnen und Schülern mit hoher Mathematikangst ( $b = -.80$ ,  $KI_{95}: -1.11/-0.49$ ). Bei Schülerinnen und Schülern mit niedriger Mathematikangst zeigte sich hingegen kein signifikanter Effekt ( $b = .17$ ,  $p = .264$ ). Die differenziertere Betrachtung anhand von Quartilen zeigte für Schülerinnen und Schüler mit hoher sowie sehr hoher Mathematikangst (3. und 4. Quartil) signifikante Effekte (3. Quartil:  $b = -0.73$ ;  $KI: -1.13/-0.32$ ; 4. Quartil:  $b = -0.95$ ;  $KI: -1.48/-0.43$ ), während sich für die beiden anderen Quartile keine signifikanten Effekte zeigen ließen. Auch Hypothese 3b konnte damit bestätigt werden.

**Tabelle 5.4: Ergebnisse der Mehrebenenanalysen zur Vorhersage der analytischen Problemlöseleistung (Studie 4b): Unstandardisierte Regressionskoeffizienten (95%-Konfidenzintervall in eckigen Klammern) und aufgeklärte Varianz**

	Modell 1b	Modell 2b	Modell 3b
Kontext (0=Problemlösen, 1=Mathematik)	-.24[-.45/-.02]*	-.28[-.49/-.06]*	-.25[-.46/-.04]*
Mathematisches Selbstkonzept (MSK; z-standardisiert)		-.00[-.16/.16]	
Interaktion: MSK*Kontext		.27[.05/.50]*	
Mathematikangst (MA; z-standardisiert)			.11[-.07/.29]
Interaktion: MA*Kontext			-.42[-.66/-.19]*
$R^2$	.01	.05	.06

Anmerkungen.  $R^2$ : Anteil der aufgeklärten Varianz auf Individualebene; \* $p < .05$ ;  $N = 259$ .

## 5.6 Diskussion

### 5.6.1 Zusammenfassung der Ergebnisse

Ziel von Studie 4 war zum einen die Prüfung der *Potenzialnutzungshypothese*, der zufolge von einer mangelnden Nutzung fachbezogener Kompetenzen durch die Schülerinnen und Schüler bei der Bearbeitung von PISA-Testaufgaben ausgegangen werden kann. Zum anderen wurde überprüft, welche Faktoren für dieses Nutzungsdefizit verantwortlich sein könnten. In Studie 4a und Studie 4b konnten alle postulierten Hypothesen bestätigt werden, wenngleich die Varianzaufklärung durch die betrachteten Prädiktoren in beiden Studien relativ gering war. Die Leistungen in den Mathematikaufgaben und die Leistungen in den analytischen Problemlöseaufgaben fielen unter der fachbezogenen mathematischen Kontextbedingung geringer aus als unter der fächerübergreifenden Problemlösebedingung (Hypothese 1a und 1b). Zusätzlich konnte ein negativer Effekt der mathematischen Kon-

textbedingung auf die Mathematikleistungen und auf die analytischen Problemlöseleistungen, insbesondere bei Schülerinnen und Schülern mit einem geringen mathematischen Selbstkonzept (Hypothese 2a und 2b) sowie bei Schülerinnen und Schülern mit hoher Mathematikangst (Hypothese 3a und 3b), belegt werden.

### 5.6.2 Theoretische und praktische Implikationen

Die Haupteffekte der kontextuellen Einbettung in Studie 4a und 4b machen deutlich, dass es sich um durchaus stabile Effekte handelt und nicht nur die Leistungen in den analytischen Problemlöseaufgaben durch die Einbettung in einen mathematischen und damit vermeintlich „falschen“ Kontext geringer werden. Vielmehr erhöhen sich auch die Leistungen in den Mathematikaufgaben durch die bloße Einbettung in einen fächerübergreifenden Kontext, obwohl zu ihrer Lösung anspruchsvolleres mathematisches Wissen notwendig ist. Dies zeigt, dass die untersuchten Schülerinnen und Schüler tatsächlich über die zur Lösung der Aufgaben notwendigen fachbezogenen Kompetenzen verfügen, sie diese aber unter der mathematischen Kontextbedingung nicht entsprechend nutzen.

Zusätzlich konnten das mathematische Selbstkonzept und die Mathematikangst als Faktoren identifiziert werden, die den Zusammenhang zwischen kontextueller Einbettung und Schülerleistungen zum Teil moderieren. So sorgen ein geringes mathematisches Selbstkonzept und/oder eine hohe Mathematikangst dafür, dass eine mathematische Kontexteinbettung sich besonders negativ auf die Schülerleistungen auswirkt. Bei Schülerinnen und Schülern mit hohem mathematischen Selbstkonzept und/oder geringer Mathematikangst kommt der Effekt der mathematischen Kontexteinbettung hingegen nicht zum Tragen. Die genauere Betrachtung zeigt, dass die Effekte nur bei Schülerinnen und Schülern in den Extrembereichen der Verteilungen (1. Quartil des mathematischen Selbstkonzepts und/oder 4. sowie bei analytischem Problemlösen 3. und 4. Quartil der Mathematikangst) auftreten. Allerdings reduziert sich durch die getrennte Betrachtung von Quartilsgruppen die Stichprobengröße deutlich, sodass diese Ergebnisse lediglich Hinweise auf etwaige Risikogruppen liefern können.



Die dargestellten Effekte liefern einen zusätzlichen Erklärungsansatz für die bei PISA 2003 festgestellte Diskrepanz zwischen der (im internationalen Vergleich hohen) Kompetenz im analytischen Problemlösen und der (im internationalen Vergleich geringeren) Mathematikkompetenz der Schülerinnen und Schüler in Deutschland (Leutner et al., 2004). Die Ergebnisse zeigen, dass – zusätzlich zur Potenzialausschöpfungshypothese – auch von einer mangelnden Nutzung tatsächlich vorhandener Fachkompetenzen in der konkreten Testsituation – im Sinne der Potenzialnutzungshypothese – ausgegangen werden kann. Dies gilt insbesondere bei Schülerinnen und Schülern mit geringem mathematischen Selbstkonzept und/oder hoher Mathematikangst.

Die Ergebnisse liefern damit einen weiteren Beleg für die Bedeutung kontextueller Einflüsse auf kognitive Leistungen (z. B. Barnett & Ceci, 2002; van den Heuvel-Panhuizen, 2005), in diesem Fall für den Bereich der fächerübergreifenden analytischen Problemlösekompetenz und der Mathematikkompetenz. So stehen die Ergebnisse im Einklang mit Zöttl et al. (2007), die beim Vergleich von *strukturgleichen* innermathematischen und außermathematischen Problemlöseaufgaben geringere Lösungsquoten für innermathematische Problemlöseaufgaben zeigen konnten. Gleichzeitig erweitert die vorliegende Untersuchung den Kenntnisstand dahingehend, dass negative Effekte einer mathematischen Kontexteinkleidung auch für *identische* Aufgaben gezeigt werden konnten. Zöttl et al. (2007) vermuten, dass ein außermathematischer Problemlösekontext vergleichsweise motivierender wirkt und dadurch, vermittelt über die Anstrengungsbereitschaft, zu besseren Leistungen führt. In der vorliegenden Untersuchung wurden keine vermittelnden Variablen untersucht, sodass auch hier über die zugrundeliegenden Wirkmechanismen nur spekuliert werden kann. Es konnte aber gezeigt werden, dass die mathematische Kontexteinkleidung insbesondere bei Personen mit einem geringen mathematischen Selbstkonzept sowie bei Personen mit einer hohen Mathematikangst Einflüsse auf die Leistungen hat. Das Setzen weniger anspruchsvoller Teilziele, eine Belastung des Arbeitsgedächtnisses durch aufgabenirrelevante Kognitionen sowie eine geringere aktuelle Motivation und Anstrengungsbereitschaft könnten somit eine wichtige Rolle bei der Vermittlung dieser Einflüsse spielen (vgl. Bouffard-Bouchard et al., 1991; Cassady & Johnson, 2002; Owens, Stevenson, Hadwin, & Norgate, 2012).

Aus den dargestellten Ergebnissen lassen sich Implikationen sowohl für die Gestaltung instruktionaler Maßnahmen zur Förderung der Mathematikkompetenz als auch zur Gestaltung von Testinstrumenten ableiten. Die hier untersuchte Potenzialnutzungshypothese ist nicht als Alternativhypothese zur Potenzialausschöpfungshypothese zu verstehen, sondern ergänzt diese. Somit sprechen die präsentierten Ergebnisse auch nicht gegen letztere. Vielmehr ist es wahrscheinlich, dass die in PISA 2003 beschriebene Diskrepanz sowohl auf nicht ausgeschöpfte kognitive Potenziale im Fachunterricht als auch auf mangelnde Nutzung vorhandener fachbezogener Kompetenzen in der Testsituation zurückzuführen ist.

Gemäß der Transferforschung ist ein positiver Transfer am ehesten bei Aufgaben mit strukturellen Gemeinsamkeiten zu erwarten (Barnett & Ceci, 2002; Klauer, 2010), wie sie für die Bereiche der analytischen Problemlösekompetenz und der Mathematikkompetenz prinzipiell gegeben zu sein scheint. Eine stärkere Berücksichtigung dieser Gemeinsamkeiten, beispielsweise mithilfe heuristischer oder prozessorientierter Lösungsbeispiele, die neben der Problemstellung und den Lösungsschritten auch den heuristischen Problemlöseprozess verdeutlichen, erscheint hier ein erfolgversprechender Ansatz zu sein (z. B. Zöttl, Ufer & Reiss, 2010). Ebenso könnten Instruktionsformen, die fächerübergreifende und fachspezifische Inhalte und Kontexte verbinden, erfolgreich sein. Dafür spricht die eingangs erwähnte Differenz in PISA 2003 zugunsten der Mathematik in den Niederlanden als einem Land, in dem diese Form der *Realistic Mathematics Education* eine besondere Tradition hat (z. B. de Lange, 1996; Hickendorff, Van Putten, Verhelst & Heiser, 2010). Darüber hinaus sollten die Gemeinsamkeiten von analytischer Problemlösekompetenz und mathematischer Kompetenz, im Sinne des mathematischen Problemlösens, möglichst explizit gemacht werden, um gerade Schülerinnen und Schüler mit besonders ungünstigen motivationalen und emotionalen Merkmalsausprägungen zu erreichen. So kann es diesen Schülerinnen und Schülern ermöglicht werden, ihre Potenziale und Kompetenzen, die in fächerübergreifenden Kontexten sichtbar werden, auch fachspezifisch anwenden zu können.

### 5.6.3 Grenzen und Ausblick

Die beiden hier dargestellten Untersuchungen ermöglichen keine gemeinsame Prüfung von Potenzialnutzungs- und Potenzialausschöpfungshypothese. Zur Untersuchung letzterer wären umfangreichere Interventionsstudien für Personen mit hohem versus niedrigem Potenzial (operationalisiert über die Differenz von analytischer Problemlöse- und Mathematikkompetenz) notwendig, die letztendlich auch international vergleichend angelegt sein sollten.

Die geringe Anzahl von Testaufgaben zur Erfassung der analytischen Problemlösekompetenz und der Mathematikkompetenz, die auch Grund für die niedrige Reliabilität beider Skalen ist, führt dazu, dass beide Kompetenzen nicht in ihrer gesamten Breite abgebildet werden konnten. Somit lassen sich differenzielle Effekte zum Beispiel für unterschiedliche mathematische Inhaltsbereiche oder unterschiedliche Arten von Problemstellungen (siehe Studie 2) nicht gänzlich ausschließen.

Für die weitere Forschung gilt es, genau hier anzusetzen und die Strukturen fächerübergreifender analytischer Problemlösekompetenz und mathematischer Problemlösekompetenz weiter aufzuklären, um sich der Antwort auf die Frage zu nähern, wie das kognitive Potenzial von Schülerinnen und Schülern zum Erwerb fachbezogener Kompetenzen optimal ausgeschöpft und/oder die konkrete Anwendung und Nutzung fachbezogener Kompetenzen erleichtert werden kann. Dies gilt gleichermaßen für die Domäne der Mathematik als auch für die Naturwissenschaften, denn auch hier gibt es Hinweise auf ungenutzte Potenziale (Rumann, Fleischer, Stawitz, Wirth & Leutner, 2010).

Interessant ist, dass bei PISA 2012, in der Problemlösen erneut erhoben wurde<sup>13</sup>, die bei PISA 2003 beobachtete große Diskrepanz zwischen analytischem Problemlösen und Mathematik verschwunden ist (OECD, 2014, Kap. 2; Sälzer, Reiss, Schiepe-Tiska, Prenzel & Heinze, 2013): Im Jahr 2012 lagen die Ergebnisse der Schülerinnen und Schüler in Deutschland in beiden Bereichen oberhalb des OECD-Durchschnitts (Problemlösen: 509 Punkte,

---

<sup>13</sup> Problemlösen wurde hier als Kombination von analytischem (*static*) und komplexem (*interactive*) Problemlösen und mit deutlichem Schwerpunkt auf dem komplexen Problemlösen erfasst (OECD, 2014).

Mathematik: 514 Punkte auf der PISA-Metrik). Inwieweit dieses Ergebnis auf eine verbesserte Potenzialnutzung zurückgeführt werden kann, sollte Gegenstand weiterer Forschung sein.

In zukünftigen Studien wäre zudem die Erweiterung des hier angewandten differenziellen Forschungsansatzes um eine kognitionspsychologische Perspektive wünschenswert (vgl. Renkl, 2012), also beispielsweise die zusätzliche Berücksichtigung von Prozessmaßen der analytischen Problemlösekompetenz und der mathematischen und naturwissenschaftlichen Kompetenzen (vgl. Naumann, Goldhammer, Rölke & Stelter, 2014). So könnten neben aggregierten Informationen über die Produkte von Lern- oder Problemlöseprozessen auch Informationen über das konkrete Vorgehen und die tatsächlich eingesetzten Strategien der Schülerinnen und Schüler gewonnen werden.

## 5.7 Literatur

- Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Barab, S. A. & Plucker, J. A. (2002). Smart people or smart contexts? Cognition, ability, and talent development in an age of situated approaches to knowing and learning. *Educational Psychologist*, 37, 165–182. doi: 10.1207/S15326985EP3703\_3
- Barnett, S. M. & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637. doi: 10.1037//0033-2909.128.4.612
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M. & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGraw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Dordrecht: Springer. doi: 10.1007/978-94-007-2324-5\_2
- Bjorklund, D. F. & Rosenblum, K. E. (2002). Context effects in children's selection and use of simple arithmetic strategies. *Journal of Cognition and Development*, 3, 225–242. doi: 10.1207/S15327647JCD0302\_5
- Blessing, S. B. & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 792–810. doi: 10.1037//0278-7393.22.3.792
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, C. H. (2004). Mathematische Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47–92). Münster: Waxmann.

- Bouffard-Bouchard, T., Parent, S. & Larivée, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. *International Journal of Behavioral Development*, 14, 153–164. doi: 10.1177/016502549101400203
- Carraher, T. N., Carraher, D. W. & Schliemann, A. D. (1985). Mathematics in the streets and in the schools. *British Journal of Developmental Psychology*, 3, 21–29. doi: 10.1111/j.2044-835X.1985.tb00951.x
- Cassady, J. C. & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27, 270–295. doi: 10.1006/ceps.2001.1094
- Chappell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A. et al. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology*, 97, 268–274. doi: 10.1037/0022-0663.97.2.268
- Clements, M. (1980). Analyzing children's errors on written mathematical tasks. *Educational Studies in Mathematics*, 11, 1–21. doi: 10.1007/BF00369157
- De Lange, J. (1996). Real problems with real world mathematics. In C. Alsina, J. M. Alvarez, M. Niss, A. Pérez, L. Rico & A. Stard (Eds.), *Proceedings of the 8th international congress on mathematical education, Sevilla July 1996* (pp. 83–110). Sevilla, Spain: Thales.
- Ferla, J., Valcke, M. & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19, 499–505. doi: 10.1016/j.lindif.2009.05.004
- Fleischer, J., Koeppen, K., Kenk, M., Klieme, E. & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. *Zeitschrift für Erziehungswissenschaft*, 16 (suppl. 1), 5–22.
- Fleischer, J., Wirth, J., Rumann, S. & Leutner, D. (2010). Strukturen fächerübergreifender und fachlicher Problemlösekompetenz – Analyse von Aufgabenprofilen. *Zeitschrift für Pädagogik*, 56. Beiheft, 239–248.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684–704.

- Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester, UK: Wiley.
- Greiff, S., Kretzschmar, A. & Leutner, D. (2014). Problemlösen in der Pädagogischen Psychologie. *Zeitschrift für Pädagogische Psychologie*, 28, 161–166. doi: 10.1024/1010-0652/a000140
- Helmke, A. & van Aken, M. A. G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, 87, 624–637. doi: 10.1037/0022-0663.87.4.624
- Hickendorff, M., Van Putten, C. M., Verhelst, N. D. & Heiser, W. J. (2010). Individual differences in strategy use on division problems: Mental versus written computation. *Journal of Educational Psychology*, 102, 438–452. doi: 10.1037/a0018177
- Klieme, E., Hartig, J. & Rauch, D. (2008). The Concept of competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Göttingen: Hogrefe.
- Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins & M. Aydede (Eds.), *The cambridge handbook of situated cognition* (pp. 264–306). Cambridge, UK: Cambridge University Press.
- Klauer, K. J. (2010). *Transfer des Lernens. Warum wir oft mehr lernen als gelehrt wird*. Stuttgart: Kohlhammer.
- Kotovsky, K., Hayes, J. R. & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive psychology*, 17, 248–294. doi: 10.1016/0010-0285(85)90009-X
- Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, UK: Cambridge University Press.
- Leutner, D., Fleischer, J., Grünkorn, J. & Klieme, E. (Eds.). (2017). *Competence assessment in education: Research, models and instruments*. Berlin: Springer. doi: 10.1007/9783-319-50030-0
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147–175). Münster: Waxmann.

- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, 30, 520–540. doi: 10.2307/749772
- Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology*, 84, 35–42. doi: 10.1037/0022-0663.84.1.35
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76, 397–416. doi: 10.1111/j.1467-8624.2005.00853.x
- Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, 26, 49–63. doi: 10.1023/A:1003088013286
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. doi: 10.1037/0003-066X.50.9.741
- Meyer, W.-U. (1984). *Das Konzept von der eigenen Begabung*. Bern: Huber.
- Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau*, 55, 19–27. doi: 10.1026/0033-3042.55.1.19
- Naumann, J., Goldhammer, F., Rölke, H. & Stelter, A. (2014). Erfolgreiches Problemlösen in technologiebasierten Umgebungen: Wechselwirkungen zwischen Interaktionsschritten und Aufgabenanforderungen. *Zeitschrift für Pädagogische Psychologie*, 28, 193–203. doi: 10.1024/1010-0652/a000134
- OECD. (2003). *The PISA 2003 assessment framework – Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing.
- OECD. (2004). *Problem solving for tomorrow's world. First measurements of cross-curricular competencies from PISA 2003*. Paris: OECD Publishing.
- OECD. (2005). *PISA 2003. Technical report*. Paris: OECD Publishing.
- OECD. (2014). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems* (Vol. 5). Paris: OECD Publishing.



- Owens, M., Stevenson, J., Hadwin, J. A. & Norgate, R. (2012). Anxiety and depression in academic performance: An exploration of the mediating factors of worry and working memory. *School Psychology International*, 33, 433–449. doi: 10.1177/0143034311427433
- Pajares, F. & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, 24, 124–139. doi: 10.1006/ceps.1998.0991
- PISA-Konsortium Deutschland (Hrsg.). (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T. & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Chicago, IL: SSI.
- Reiss, K. & Törner, G. (2007). Problem solving in the mathematics classroom: The German perspective. *ZDM – The International Journal on Mathematics Education*, 39, 431–441. doi: 10.1007/s11858-007-0040-5
- Renkl, A. (2012). Modellierung von Kompetenzen oder von interindividuellen Kompetenzunterschieden: Ein unterschätzter Unterschied. *Psychologische Rundschau*, 63, 50–53. doi: 10.1026/0033-3042/a000110
- Rost, D. H. & Sparfeldt, J. R. (2002). Facetten des schulischen Selbstkonzepts: Ein Verfahren zur Messung des differentiellen Selbstkonzepts schulischer Leistungen und Fähigkeiten (DISK-Gitter). *Diagnostica*, 48, 130–140. doi: 10.1026//0012-1924.48.3.130
- Rumann, S., Fleischer, J., Stawitz, H., Wirth, J. & Leutner, D. (2010). Vergleich von Profilen der Naturwissenschafts- und Problemlöse-Aufgaben der PISA 2003-Studie. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 315–327.
- Sälzer, C., Reiss, K., Schiepe-Tiska, A., Prenzel, M. & Heinze, A. (2013). Zwischen Grundlagenwissen und Anwendungsbezug: Mathematische Kompetenz im internationalen Vergleich. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 47–97). Münster: Waxmann.
- Scherer, R. & Gustafsson, J.-E. (2015). The relations among openness, perseverance, and performance in creative problem solving: A substantive-methodological approach. *Thinking Skills and Creativity*, 18, 4–17. doi: 10.1016/j.tsc.2015.04.004

- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.
- Schwarzer, R. (2000). *Streß, Angst und Handlungsregulation* (4. überarb. Aufl.). Stuttgart: Kohlhammer.
- Stankov, L., Lee, J., Luo, W. & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, 22, 747–758. doi: 10.1016/j.lindif.2012.05.013
- Valentine, J. C., DuBois, D. L. & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39, 111–133. doi: 10.1207/s15326985ep3902\_3
- van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics*, 25(2), 2–9.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Seattle, WA: Hogrefe.
- Wang, M.-T. & Eccles, J. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction*, 28, 12–23. doi: 10.1016/j.learninstruc.2013.04.002
- Zöttl, L., Heinze, A. & Reiss, K. (2007). Problemlösen im Kontext – Unterschiede in der Bearbeitung von Alltagsproblemen und mathematischen Problemen. In A. Peter-Koop & A. Bikner-Ahsbahr (Hrsg.), *Mathematische Bildung – Mathematische Leistung. Festschrift für Michael Neubrand zum 60. Geburtstag* (S. 217–232). Hildesheim: Franzbecker.
- Zöttl, L., Ufer, S. & Reiss, K. (2010). Modelling with heuristic worked examples in the KOMMA learning environment. *Journal für Mathematik-Didaktik*, 31, 143–165. doi: 10.1007/s13138-010-0008-9

## 6 GESAMTDISKUSSION

Ausgangspunkt dieser Arbeit waren Ergebnisse der PISA-Studie 2003, die darauf hindeuten, dass sich beim fächerübergreifenden analytischen Problemlösen kognitives Potenzial manifestiert, welches in der Schule nicht hinreichend zum Aufbau fachbezogener Kompetenzen, insbesondere in der Mathematik, aber auch in den Naturwissenschaften, genutzt wird (Potenzialausschöpfungshypothese; Leutner, Klieme, Meyer & Wirth, 2004; OECD, 2004, Kap. 4; Rost, Walter, Carstensen, Senkbeil & Prenzel, 2004). Um dieses Potenzial für didaktische Maßnahmen zum Aufbau und zur Förderung fachbezogener Kompetenzen nutzbar machen zu können, müssen zentrale Aspekte der Validität der Modellierung analytischer Problemlösekompetenz umfassender untersucht werden, als dies in der bisherigen Forschung geschehen ist. Diese Aspekte beziehen sich unter anderem auf Fragen der Binnenstruktur, der Abgrenzbarkeit von anderen schulleistungsrelevanten Konstrukten sowie auf Fragen der prädiktiven Kraft der analytischen Problemlösekompetenz für fachbezogene Kompetenzen.

Ziel der Arbeit war es daher erstens, verschiedene Aspekte der Validität der Modellierung der analytischen Problemlösekompetenz zu prüfen und zweitens, eine ergänzende Erklärung für die Potenzialausschöpfungshypothese sowie die Bedeutung motivationaler und emotionaler Faktoren für erfolgreiches Problemlösen zu untersuchen. Hierzu wurden in den Studien 1 bis 3 Fragen der faktoriellen, diskriminanten, prognostischen und inkrementellen Validität als Aspekte der Konstruktvalidität (sensu Messick, 1995; siehe auch AERA, APA & NCME, 1999 sowie Hartig, Frey & Jude, 2012) der Modellierung analytischer Problemlösekompetenz untersucht. Dabei wurde auf verschiedene Datensätze der PISA-Studie 2003 zurückgegriffen, in der Kompetenzen als „prinzipiell erlernbare, mehr oder minder bereichsspezifische Kenntnisse, Fertigkeiten und Strategien“ definiert wurden (Baumert, Stanat & Demmrich, 2001, S. 22). Motivationale und emotionale Faktoren wurden in dieser Definition, dem üblichen Vorgehen im Rahmen der Forschung zur Modellierung von Kompetenzen entsprechend, vor allem aus pragmatischen Gründen ausgeschlossen (Fleischer, Koeppen, Kenk, Klieme & Leutner, 2013; Koeppen, Hartig, Klieme & Leutner, 2008; Leutner, Fleischer, Grünkorn & Klieme, 2017). In Studie 4 wurde diese Perspektive

erweitert, indem zum einen geprüft wurde, ob in Ergänzung zur Potenzialausschöpfungshypothese auch von einer mangelnden Nutzung vorhandener fachbezogener Kompetenzen durch die Schülerinnen und Schüler bei der Bearbeitung von PISA-Testaufgaben ausgegangen werden kann (Nutzungsdefizit). Zum anderen wurden motivationale und emotionale Faktoren untersucht, die dafür verantwortlich sein könnten, dass Schülerinnen und Schüler ihr kognitives Potenzial und ihre Kompetenzen bei der Bearbeitung von Testaufgaben nicht erfolgreich nutzen.

## 6.1 Zentrale Ergebnisse

Im Folgenden werden die zentralen Ergebnisse bezogen auf die Hypothesen und Forschungsfragen der einzelnen Studien dargestellt.

*In Studie 1 wurde davon ausgegangen, dass eine dreidimensionale Modellierung analytischer Problemlösekompetenz entsprechend den drei Arten von Problemstellungen (1) „Entscheidungen treffen“, (2) „Systeme analysieren und entwerfen“ und (3) „Fehler suchen“ die Datenstruktur in PISA 2003 adäquater abbildet als eine eindimensionale Modellierung ohne entsprechende Differenzierung.*

*Für die Dimensionen des dreidimensionalen Modells sollten sich darüber hinaus differenzielle Zusammenhänge mit externen Variablen zeigen lassen.*

Als zentrales Ergebnis der Studie 1 ließen sich die drei Arten von Problemstellungen zur Erfassung analytischer Problemlösekompetenz empirisch trennen, und eine entsprechende dreidimensionale Modellierung der Daten zeigte eine bessere Modellpassung als ein eindimensionales Modell. Es ergaben sich darüber hinaus Hinweise auf differenzielle Zusammenhänge der Dimensionen des dreidimensionalen Modelles mit externen Variablen, allerdings ließen sich diese nicht für alle drei Dimensionen und nicht für alle betrachteten externen Variablen inferenzstatistisch absichern.

*In Studie 2 wurde davon ausgegangen, dass ein Nested-Faktor-Modell, welches von der multiplen Bedingtheit analytischer Problemlöseleistungen durch analytische Problemlösekompetenz und fluide Fähigkeiten ausgeht, eine bessere Modellpassung aufweist als das Standardmodell, welches analytische Problemlöseleistungen und Leistungen in Items zur Erfassung fluider Fähigkeiten auf jeweils unterschiedliche Faktoren zurückführt.*

*Der um fluide Fähigkeiten bereinigte Faktor der analytischen Problemlösekompetenz im Nested-Faktor-Modell sollte darüber hinaus signifikante Varianz und substantielle Faktorladungen aufweisen.*

*Ferner sollten sich für diesen Faktor, im Vergleich zum entsprechenden Faktor im Standardmodell, differenzielle Zusammenhänge mit externen Variablen zeigen.*

Als zentrales Ergebnis der Studie 2 zeigte das untersuchte Nested-Faktor-Modell eine bessere Modellpassung gegenüber dem Standardmodell und dem Einfaktormodell. Zudem ließen sich für den Faktor der analytischen Problemlösekompetenz im Nested-Faktor-Modell eine signifikante Varianz sowie substantielle Faktorladungen schätzen. Darüber hinaus zeigten sich für diesen Faktor, im Vergleich zum entsprechenden Faktor der analytischen Problemlösekompetenz im Standardmodell, signifikant unterschiedliche Zusammenhänge mit fachbezogenen Kompetenzen.

*In Studie 3 wurde davon ausgegangen, dass die analytische Problemlösekompetenz zukünftige fachbezogene Kompetenzen in der Mathematik und den Naturwissenschaften vorherzusagt.*

*Es wurde gefragt, inwiefern die analytische Problemlösekompetenz geeignet ist, auch die Kompetenzentwicklung (Vorhersage zukünftiger fachbezogener Kompetenzen unter Kontrolle der jeweiligen Ausgangskompetenzen) in der Mathematik und den Naturwissenschaften vorherzusagen.*

*Darüber hinaus wurde gefragt, inwiefern die analytische Problemlösekompetenz geeignet ist, die Kompetenzentwicklung in der Mathematik und den Naturwissenschaften auch unter Kontrolle der fluiden Fähigkeiten vorherzusagen.*

*Ferner wurde davon ausgegangen, dass die analytische Problemlösekompetenz, über den Effekt der fluiden Fähigkeiten hinaus, substanziell inkrementelle Varianz zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften aufklärt.*

*Es wurde gefragt, inwiefern die analytische Problemlösekompetenz, über die Effekte der jeweiligen Ausgangskompetenzen hinaus, substanziell inkrementelle Varianz zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften aufklärt.*

*Darüber hinaus wurde gefragt, inwiefern die analytische Problemlösekompetenz, über den Effekt der fluiden Fähigkeiten hinaus, substanziell inkrementelle Varianz der Kompetenzentwicklung in der Mathematik und in den Naturwissenschaften aufklärt.*

Als zentrales Ergebnis der Studie 3 zeigten sich signifikante Effekte der analytischen Problemlösekompetenz sowohl bei der Vorhersage zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften als auch bei der Vorhersage der Kompetenzentwicklung (Vorhersage zukünftiger fachbezogener Kompetenzen unter Kontrolle der jeweiligen Ausgangskompetenzen) in diesen Domänen, jeweils auch unter Kontrolle der fluiden Fähigkeiten. Dabei wies die analytische Problemlösekompetenz zwar ein substanzielles Inkrement gegenüber den fluiden Fähigkeiten auf, nicht aber gegenüber den jeweiligen Ausgangskompetenzen bei der Vorhersage zukünftiger fachbezogener Kompetenzen. Ebenso zeigte sich bei der Vorhersage der Kompetenzentwicklung kein substanzielles Inkrement der analytischen Problemlösekompetenz gegenüber den fluiden Fähigkeiten.

*In Studie 4 wurde davon ausgegangen, dass Schülerleistungen sowohl in Mathematikaufgaben als auch in Problemlöseaufgaben aus PISA 2003 geringer ausfallen, wenn diese in einen Mathematikkontext eingebettet sind im Vergleich zu einer Einbettung in einen alltäglichen Problemlösekontext.*

*Es wurde ferner erwartet, dass der Effekt der kontextuellen Einkleidung auf die Schülerleistungen besonders ausgeprägt ist bei Schülerinnen und Schülern mit geringem mathematischem Selbstkonzept sowie bei Schülerinnen und Schülern mit hoher Mathematikangst.*

Als zentrales Ergebnis der Studie 4 konnten sowohl bei den Mathematik- als auch bei den Problemlöseaufgaben die postulierten Effekte der kontextuellen Einkleidung aufgezeigt werden. So fielen die Schülerleistungen in den Mathematikaufgaben und in den analytischen Problemlöseaufgaben unter der fachbezogenen mathematischen Kontextbedingung geringer aus als unter der fächerübergreifenden Problemlösebedingung. Ebenso konnte ein negativer Effekt der mathematischen Kontextbedingung auf die Leistungen in den Mathematikaufgaben und in den analytischen Problemlöseaufgaben, insbesondere bei Schülerinnen und Schülern mit einem geringen mathematischen Selbstkonzept sowie bei Schülerinnen und Schülern mit hoher Mathematikangst, nachgewiesen werden.

## 6.2 Theoretischer Ertrag

### 6.2.1 Faktorielle Validität

Hinsichtlich der Binnenstruktur des bei PISA 2003 eingesetzten Testinstruments zur Erfassung analytischer Problemlösekompetenz lässt sich festhalten, dass die dabei verwendeten Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen* unterschiedliche Aspekte der analytischen Problemlösekompetenz erfassen, die empirisch trennbare Dimensionen darstellen, hinsichtlich derer sich Personen unterscheiden können. Eine besondere Nähe zeigt sich hierbei für die beiden Dimensionen *Entscheidungen treffen* und *Systeme analysieren und entwerfen*, die sehr hoch korreliert sind, und deren Zusammenhänge mit fachbezogenen Kompetenzen in der Mathematik, den Naturwissenschaften und im Lesen sowie mit fluiden Fähigkeiten in einer ähnlichen Größenordnung liegen. Demgegenüber zeigen sich für die Dimension *Fehler suchen* niedrigere Zusammenhänge mit diesen Variablen, auch wenn sich diese Unterschiede (außer bei der Lesekompetenz) nur für den Vergleich der Dimensionen *Fehler suchen* und *Systeme analysieren und entwerfen* inferenzstatistisch absichern lassen. Problemstellungen der Art *Entscheidungen treffen* und *Systeme analysieren und entwerfen* ähneln den Projektaufgaben, wie sie in PISA 2003 eingesetzt wurden, bei denen ein Projekt in einzelne Arbeitsschritte gegliedert wird, und zu denen dann einzelne Testaufgaben gestellt werden (Klieme, Hartig & Wirth, 2005). Diese Ähnlichkeit in der Konzeption beider Arten

von Problemstellungen mag das ähnliche Befundmuster der beiden Dimensionen *Entscheidungen treffen* und *Systeme analysieren und entwerfen* erklären.

Die Ergebnisse von Studie 1 sprechen für die Ausdifferenzierung der analytischen Problemlösekompetenz im Sinne einer mehrdimensionalen Struktur und damit einerseits gegen die im Rahmen von PISA vorgenommene eindimensionale Skalierung (OECD, 2005). Andererseits scheint der Informationsverlust bei einer eindimensionalen Skalierung der analytischen Problemlösekompetenz nicht so gravierend zu sein, dass er der intendierten Verwendung der Testergebnisse, nämlich der Bereitstellung von Informationen für ein Bildungsmonitoring auf Systemebene, entgegenenspräche.

### 6.2.2 Diskriminante Validität

Fluide Fähigkeiten als zentrale Sekundärfähigkeiten der allgemeinen kognitiven Fähigkeiten (Carroll, 1993) und damit eines der zentralen Intelligenzkonzepte (Horn & Noll, 1997) weisen eine deutliche konzeptuelle Nähe zur analytischen Problemlösekompetenz auf (z. B. Leutner, Funke, Klieme & Wirth, 2005). Dennoch gibt es Belege für die empirische Trennbarkeit beider Konstrukte (z. B. Leutner, Wirth, Klieme & Funke, 2005; Leutner et al., 2004; Scherer & Tiemann, 2014). Bisherige Untersuchungen haben jedoch zum einen analytische Problemlösekompetenz nicht als mehrdimensionales Konstrukt modelliert, bei dem unterschiedliche Arten von Problemstellungen empirisch trennbare Dimensionen abbilden, die unterschiedlich stark durch fluide Fähigkeiten beeinflusst werden. Zum anderen wurde die theoretische Annahme der multiplen Bedingtheit von Leistungen in analytischen Problemstellungen durch analytische Problemlösekompetenz und fluide Fähigkeiten bisher nicht berücksichtigt. Das in Studie 2 favorisierte Nested-Faktor-Modell (Gustafsson & Balke, 1993) leistet genau dies. Es stellt zum einen die theoretisch adäquatere Modellierung der Zusammenhänge von analytischer Problemlösekompetenz und fluiden Fähigkeiten dar, und bildet zum anderen die Datenstruktur besser ab als ein Standardmodell, welches analytische Problemlöseleistungen und Leistungen in Items zur Erfassung fluider Fähigkeiten auf unterschiedliche Faktoren zurückführt. Fluide Fähigkeiten reichen demnach nicht aus, um die Leistungen in analytischen Problemlöseaufgaben zu erklären. Hinzukommen muss



eine von fluiden Fähigkeiten unabhängige Komponente analytischer Problemlösekompetenz, welche die Leistungen in den drei Arten von Problemlöseaufgaben in etwa gleichem Maße bedingt. Demgegenüber scheinen fluide Fähigkeiten, je nach Art der zu bearbeitenden Problemstellung, einen unterschiedlich starken Einfluss auf die Qualität der Problemlösung zu nehmen.

### 6.2.3 Prognostische und inkrementelle Validität

Bisherige Untersuchungen haben Belege für den Zusammenhang analytischer Problemlösekompetenz mit fachbezogenen Kompetenzen, im Sinne der Übereinstimmungsvalidität geliefert (z. B. Leutner et al., 2004; OECD, 2005). Die analytische Problemlösekompetenz leistet darüber hinaus auch einen wichtigen Beitrag sowohl zur Vorhersage zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften als auch zur Vorhersage der Kompetenzentwicklung in diesen Domänen. Dabei sind die Effekte der analytischen Problemlösekompetenz bei der Vorhersage der Kompetenzentwicklung mit denen der fluiden Fähigkeiten vergleichbar, während die analytische Problemlösekompetenz bei der Vorhersage zukünftiger fachbezogener Kompetenzen im Vergleich zu fluiden Fähigkeiten sogar den erklärungskräftigeren Prädiktor darstellt.

Zudem klärt die analytische Problemlösekompetenz, gegenüber den fluiden Fähigkeiten, substantiell inkrementelle Varianz zukünftiger fachbezogener Kompetenzen in der Mathematik und den Naturwissenschaften auf. Umgekehrt zeigt sich für die fluiden Fähigkeiten hingegen kein entsprechendes Inkrement gegenüber der analytischen Problemlösekompetenz bei der Vorhersage zukünftiger Kompetenzen in der Mathematik und den Naturwissenschaften – ein Befund, der die besondere Rolle der analytischen Problemlösekompetenz für die Vorhersage zukünftiger fachbezogener Kompetenzen noch unterstreicht. In einer querschnittlichen Perspektive, konnten entsprechende Ergebnisse bereits bei der Modellierung der analytischen Problemlösekompetenz mithilfe des Nested-Faktor-Modells demonstriert werden, in dem sich für einen von fluiden Fähigkeiten unabhängigen Faktor der analytischen Problemlösekompetenz substantielle Korrelationen mit der Mathematikkompetenz und der naturwissenschaftlichen Kompetenz zeigen ließen. Bei der

Vorhersage zukünftiger Kompetenzen in der Mathematik und den Naturwissenschaften liefert die analytische Problemlösekompetenz hingegen keinen substanziellen *inkrementellen* Beitrag gegenüber den jeweiligen Ausgangskompetenzen, was einmal mehr die zentrale Bedeutung bereichsspezifischen Vorwissens für zukünftige Kompetenzen deutlich macht (z. B. Ackerman & Beier, 2006; Köller & Baumert, 2008). Die Bedeutung der analytischen Problemlösekompetenz für die Kompetenzentwicklung wird dadurch jedoch nicht in Abrede gestellt, allenfalls geschmälert, zumal sich für die fluiden Fähigkeiten ein analoger Befund zeigt. Es ist vielmehr davon auszugehen, dass sich – in den in Studie 3 zugrunde gelegten Daten – ein nicht unerheblicher Teil des Einflusses der analytischen Problemlösekompetenz auf die zukünftigen fachbezogenen Kompetenzen bereits zum ersten Messzeitpunkt manifestiert hat und somit, im Sinne eines indirekten Effekts, über die jeweiligen Ausgangskompetenzen auf zukünftige fachbezogene Kompetenzen vermittelt wird. Dies lässt sich jedoch auf Basis der verwendeten Datengrundlage nicht abschließend prüfen.

Zusammenfassend lassen die Ergebnisse der Studien 2 und 3 den Schluss zu, dass die analytische Problemlösekompetenz einen wichtigen Beitrag zum Aufbau fachbezogener Kompetenzen in der Mathematik und in den Naturwissenschaften zu leisten vermag.

#### **6.2.4 Potenzialnutzung und die Bedeutung motivationaler und emotionaler Faktoren**

Die Befunde von Studie 4 zur *Potenzialnutzungshypothese* zeigen, dass – in Ergänzung zur *Potenzialausschöpfungshypothese* (Leutner et al., 2004; OECD, 2004, Kap. 4) – von einer mangelnden Nutzung tatsächlich vorhandener fachbezogener Kompetenzen durch die Schülerinnen und Schüler bei der Bearbeitung von PISA-Testaufgaben ausgegangen werden kann. Damit kann ein weiterer Beleg für die Bedeutung kontextueller Einflüsse auf kognitive Leistungen (z. B. Barab & Plucker, 2002; Barnett & Ceci, 2002; van den Heuvel-Panhuizen, 2005) für den Bereich der Mathematikkompetenz und insbesondere für den Bereich der analytischen Problemlösekompetenz erbracht werden. Es sind also nicht allein die zur Lösung einer analytischen Problemlöseaufgabe notwendigen Kompetenzen, die das Lösungsverhalten in diesen Aufgaben bedingen. Auch der Kontext, in den eine Aufgabe be-

ziehungsweise Problemstellung eingebunden ist, hat Einfluss auf die Qualität der Aufgaben- beziehungsweise Problemlösung. Zöttl, Heinze und Reiss (2007) konnten bereits zeigen, dass innermathematische Problemlöseaufgaben geringere Lösungsquoten aufweisen als *strukturgleiche* außermathematische Problemlöseaufgaben. Entsprechende negative Effekte einer mathematischen Kontexteinkleidung konnten in Studie 4 dieser Arbeit auch für *identische* Aufgaben nachgewiesen werden. Über die genauen Wirkmechanismen können dabei keine definitiven Aussagen getroffen werden, da keine potenziell vermittelnden Variablen oder Prozessmaße erhoben wurden. Es konnte aber gezeigt werden, dass die mathematische Kontexteinkleidung, insbesondere bei Personen mit einem geringen mathematischen Selbstkonzept sowie bei Personen mit einer hohen Mathematikangst, negative Einflüsse auf die Leistungen ausübt. Das Setzen weniger anspruchsvoller Teilziele, eine Belastung des Arbeitsgedächtnisses durch aufgabenirrelevante Kognitionen sowie eine geringere aktuelle Motivation und Anstrengungsbereitschaft könnten somit eine wichtige Rolle bei der Vermittlung dieser Einflüsse spielen (vgl. Bouffard-Bouchard et al., 1991; Cassady & Johnson, 2002; Owens, Stevenson, Hadwin, & Norgate, 2012).

### 6.3 Praktischer Ertrag

Die in dieser Arbeit präsentierten Ergebnisse verdeutlichen, dass die Berücksichtigung unterschiedlicher Arten von analytischen Problemstellungen sowie unterschiedlicher Arten der Modellierung der Zusammenhänge von analytischer Problemlösekompetenz und fluiden Fähigkeiten unterschiedliche Ergebnisse hinsichtlich des Zusammenhangs mit externen Variablen liefern können. Insbesondere der erstgenannte Aspekt sollte bei der Itementwicklung im Rahmen der Konstruktion von neuen Testinstrumenten zur Erfassung der analytischen Problemlösekompetenz berücksichtigt werden. Für die diagnostische Praxis im Rahmen der Kompetenzdiagnostik als Spezialfall der pädagogisch-psychologischen Diagnostik, bei der es um die Vorbereitung von Entscheidungen auf Einzelfallebene geht (Klieme & Leutner 2006), könnte sich so die Möglichkeit bieten, zu differenzierteren Aussagen über spezifische Stärken und Schwächen von Schülerinnen und Schülern zu kommen. Die berichteten Ergebnisse zur inkrementellen Validität unterstreichen die generelle Bedeutung der analytischen Problemlösekompetenz für die individualdiagnostische Praxis.

So zeigen sie den Mehrwert der analytischen Problemlösekompetenz gegenüber fluiden Fähigkeiten bei der Vorhersage fachbezogener Kompetenzen sowohl in der Mathematik als auch in den Naturwissenschaften. Dies berechtigt zu der Annahme, dass durch die zusätzliche Erfassung der analytischen Problemlösekompetenz im Rahmen des diagnostischen Prozesses, die Qualität von Vorhersagen zur Vorbereitung von Entscheidungen auf Einzel-fallebene verbessert werden kann.

Auch wenn es um die Frage geht, an welchen Stellen didaktisch angesetzt werden kann, um das kognitive Potenzial, welches sich gemäß der Potenzialausschöpfungshypothese beim analytischen Problemlösen manifestiert, zum Aufbau fachbezogener Kompetenzen zu nutzen, haben diese Befunde eine praktische Relevanz. Entsprechend sollten sie bei der Konstruktion von Trainingsprogrammen und instruktionalen Maßnahmen zur Förderung analytischer Problemlösekompetenz (z. B. Buchwald, Fleischer & Leutner, 2015; Buchwald, Fleischer, Rumann, Wirth & Leutner, 2017) berücksichtigt werden, da angenommen werden kann, dass diese Fördermaßnahmen unterschiedlich wirksam sein können, je nachdem auf welche Arten von Problemstellungen sie fokussieren.

Die dargestellten Befunde zur Potenzialnutzungshypothese sprechen für eine mangelnde Nutzung vorhandener fachbezogener Kompetenzen in einer konkreten Testsituation. Die höheren Leistungen im fächerübergreifenden Kontext im Vergleich zum mathematischen Kontext, lassen die Vermutung zu, dass Instruktionsformen, die fächerübergreifende und fachspezifische Inhalte und Kontexte verbinden, besonders erfolgreich sein können. Dafür sprechen die PISA 2003-Ergebnisse der Schülerinnen und Schüler aus den Niederlanden, als einem Land, in dem diese Form der *Realistic Mathematics Education* eine besondere Tradition hat (z. B. de Lange, 1996; Hickendorff, Van Putten, Verhelst & Heiser, 2010). Im Vergleich zu Deutschland zeigte sich demgegenüber eine Differenz zwischen der Mathematikkompetenz und der analytischen Problemlösekompetenz zugunsten der Mathematik (OECD, 2004, Kap. 4). Ferner könnten gerade Schülerinnen und Schüler mit besonders ungünstigen motivationalen und emotionalen Merkmalsausprägungen davon profitieren, wenn in entsprechenden Lern- und Testsituationen die Gemeinsamkeiten des fächerübergreifenden analytischen und des mathematischen Problemlösens (siehe die Ausführungen in den Abschnitten 1.2 und 1.3.2) möglichst explizit gemacht werden. So könnte

es diesen Schülerinnen und Schülern ermöglicht werden, ihre Potenziale und Kompetenzen, die sie in fächerübergreifenden Kontexten zum Ausdruck bringen, auch in fachspezifischen Kontexten zeigen zu demonstrieren.

## 6.4 Limitationen und Ausblick

In der Gesamtschau der Ergebnisse ergeben sich insbesondere folgende Limitationen der dargestellten Studien und daraus abgeleitete Implikationen für zukünftige Forschungsarbeiten:

Eine erste Einschränkung bezieht sich auf die zur Operationalisierung der analytischen Problemlösekompetenz verwendeten Arten von Problemstellungen. So wären unter Verwendung anderer Arten von Problemstellungen gegebenenfalls andere als die in dieser Arbeit berichteten Ergebnisse hinsichtlich der Binnenstruktur zu erwarten. Beispielsweise weisen Problemlöseaufgaben, die nach dem Prinzip von Projektaufgaben konstruiert wurden, eine größere Nähe zueinander auf, als Problemstellungen, die nach anderen Prinzipien konstruiert wurden. So wäre in zukünftigen Forschungsarbeiten zu prüfen, ob und inwiefern die *fächerverbindenden* Problemstellungen aus PISA 2003 sowie Problemstellungen von der Art, wie sie in der nationalen Ergänzungsstudie zu PISA 2000 eingesetzt wurden (Klieme et al., 2005)), und die als *fächerunabhängige* Problemstellungen bezeichnet werden können, tatsächlich dasselbe Konstrukt abbilden. Des Weiteren gilt es zu untersuchen, wie viele Dimensionen zur adäquaten Modellierung der analytischen Problemlösekompetenz in diesem Fall notwendig wären.

Eine weitere Einschränkung, die sich ebenfalls auf den Aspekt der Operationalisierung bezieht, aber weniger die theoretische, sondern vielmehr die praktische Bedeutsamkeit betrifft, bezieht sich auf die Erfassung einer von fluiden Fähigkeiten unabhängigen analytischen Problemlösekompetenz. Diese ist mit den im Rahmen von PISA eingesetzten Problemstellungen nicht möglich, sodass sich hier ein weiteres Forschungsdesiderat ergibt. Derartige alternative Problemstellungen würden die Möglichkeit bieten, das Inkrement der analytischen Problemlösekompetenz faktoriell rein und damit direkt zu erfassen. Dies würde, mit anderen Worten, die Option einer effizienten und ökonomischen Erfassung nur

derjenigen Aspekte der analytischen Problemlösekompetenz ermöglichen, die nicht bereits in Instrumenten zur Erfassung fluider Fähigkeiten miterhoben werden.

Eine zusätzliche Einschränkung betrifft die in Studie 3 verwendete Datengrundlage, die fachbezogene Kompetenzwerte von zwei Messzeitpunkten über den Verlauf eines Jahres bereitstellt. Die in Abschnitt 4.2.2 referierten Ergebnisse (Ferrer et al., 2007; Helmke & Weinert, 1997; siehe auch Stern, 2001, 2003), lassen den Schluss zu, dass die Bedeutung früherer fachbezogener Kompetenzen für die Vorhersage zukünftiger fachbezogener Kompetenzen über den zeitlichen Verlauf größer wird, während die Bedeutung von Intelligenzmaßen im Verlauf der Lerngeschichte geringer wird. Dieser Umstand könnte für die analytische Problemlösekompetenz, als ebenfalls fächerübergreifendem Konstrukt, in ähnlicher Art und Weise zutreffen. Um diese Annahme prüfen zu können, müssten entsprechende Längsschnittanalysen über eine größere Zeitspanne mit mindestens drei Messzeitpunkten durchgeführt werden. Dies würde darüber hinaus zum einen die Prüfung etwaiger Mediationseffekte der analytischen Problemlösekompetenz über fachbezogene Kompetenzen zu einem bestimmten Zeitpunkt auf zeitlich noch später erfasste fachbezogene Kompetenzen ermöglichen. Zum anderen böte es die Möglichkeit, hinreichend große Item- und Personenstichproben vorausgesetzt, elaboriertere Analyseverfahren wie Wachstumskurvenmodelle oder Latent-Change-Modelle einzusetzen (z. B. McArdle & Grimm, 2010; Schmiedek & Wolff, 2010).

Da die Abgrenzung der analytischen Problemlösekompetenz von fluiden Fähigkeiten einen zentralen Aspekt von Studie 2 und 3 darstellt, soll im Folgenden abschließend auf einen Aspekt des Geltungsbereichs der in diesen Studien dargestellten Ergebnisse eingegangen werden. Wie in Studie 2 dargestellt, werden figurale Analogieaufgaben, wie die in der vorliegenden Arbeit verwendete Subskala *Figurenanalogien des Kognitiven Fähigkeits-tests* (KFT 5-12+R, Heller & Perleth, 2000) häufig zur Erfassung des schlussfolgernden Denkens (*reasoning*) eingesetzt, welches den zentralen Primärfaktor der fluiden Fähigkeiten darstellt (Carroll, 1993). Die fluiden Fähigkeiten gelten wiederum als der zentrale Sekundärfaktor der allgemeinen kognitiven Fähigkeiten (Carroll, 1993) beziehungsweise als guter Indikator für den *g*-Faktor der Intelligenz (Carlstedt, 2001; Gustafsson, 1988). Schlussfolgerndes Denken und fluide Fähigkeiten können demnach als zentrale Intelligenzkonzepte betrachtet werden (z. B. Carroll, 1993; Carlstedt, 2001; Horn & Noll, 1997; Rost, 2009, Kap.

2). Für eine adäquate Erfassung des *g*-Faktors reicht jedoch eine einzelne Subskala eines Intelligenztests nicht aus. Nach Jensen und Weng (1994) sollten für eine hinreichend gute Indikatorisierung des *g*-Faktors mindestens zehn Subtests eingesetzt werden. Bei einer Operationalisierung allgemeiner kognitiver Grundfähigkeiten durch mehrere Subtests zur Erfassung unterschiedlicher Primärfaktoren würden sich somit wahrscheinlich stärkere Effekte der allgemeinen kognitiven Grundfähigkeiten quasi zulasten der analytischen Problemlösekompetenz bei der Vorhersage fachbezogener Kompetenzen zeigen. Ergebnisse zur inkrementellen Validität von Maßen zur Erfassung komplexer Problemlösekompetenz gegenüber einer breiten Indikatorisierung des *g*-Faktors bei der Vorhersage von Schulleistungen lassen dies vermuten (z. B. Kretzschmar, Neubert, Wüstenberg & Greiff, 2016; Lotz, Sparfeldt & Greiff, 2016).

Auf der anderen Seite kann auch für die Modellierung der analytischen Problemlösekompetenz in den Studien 1 bis 3 von einer relativ schwachen Indikatorisierung ausgegangen werden. Wie in Abschnitt 1.3.3 dargestellt, erfasst der Problemlösetests aus PISA 2003, der die Datengrundlage für diese Studien liefert, mit 18 Items die drei Arten von Problemstellungen (1) *Entscheidungen treffen*, (2) *Systeme analysieren und entwerfen* und (3) *Fehler suchen*. Es kann somit ebenfalls vermutet werden, dass eine breitere Operationalisierung der analytischen Problemlösekompetenz mit einem größeren Itempool und dem Einbezug weiterer Arten von Problemstellungen zu stärkeren Effekten bei der Vorhersage fachbezogener Kompetenzen und auch zu einem deutlicheren Inkrement der analytischen Problemlösekompetenz gegenüber fluiden Fähigkeiten geführt hätte.

Festzuhalten bleibt, dass sich die in Studie 2 und 3 der vorliegenden Arbeit präsentierten Ergebnisse auf schlussfolgerndes Denken beziehungsweise fluide Fähigkeiten als zentrale Konzepte der allgemeinen kognitiven Grundfähigkeiten beziehen. Es soll nicht der Anspruch erhoben werden, Aussagen über den *g*-Faktor allgemeiner kognitiver Fähigkeiten treffen zu können. Die Beantwortung der Frage, ob für analytische Problemlösekompetenz auch ein Mehrwert gegenüber dem *g*-Faktor bei der Vorhersage fachbezogener Kompetenzen gezeigt werden kann, bleibt zukünftigen Forschungsarbeiten vorbehalten. Die in dieser Arbeit präsentierten Ergebnisse, verbunden mit der Möglichkeit einer breiteren Indikatorisierung der analytischen Problemlösekompetenz über zusätzliche Arten von Problemstellungen, lassen dies aber zumindest nicht unwahrscheinlich erscheinen. Die Tatsache, dass

für komplexe Problemlösekompetenz, für die im Vergleich zur analytischen Problemlösekompetenz von einer geringeren Affinität zu fachbezogenen Kompetenzen ausgegangen werden kann (Fischer, Greiff, Wüstenberg, Fleischer, Buchwald & Funke 2015; Leutner, Fleischer, Wirth, Greiff & Funke, 2012), ein Inkrement gegenüber dem  $g$ -Faktor bei der Vorhersage mathematischer Kompetenzen gezeigt werden konnte (Lotz, Sparfeldt & Greiff, 2016), sprechen ebenfalls dafür.



## 6.5 Literatur

- Ackerman, P. L. & Beier, M. E. (2006). Determinants of domain knowledge and independent study learning in an adult sample. *Journal of Educational Psychology, 98*, 366–381. doi: 10.1037/0022-0663.98.2.366
- AERA., APA. & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: APA.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Barab, S. A. & Plucker, J. A. (2002). Smart people or smart contexts? Cognition, ability, and talent development in an age of situated approaches to knowing and learning. *Educational Psychologist, 37*, 165–182. doi: 10.1207/S15326985EP3703\_3
- Barnett, S. M. & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612–637. doi: 10.1037//0033-2909.128.4.612
- Bouffard-Bouchard, T., Parent, S. & Larivée, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. *International Journal of Behavioral Development, 14*, 153–164. doi: 10.1177/016502549101400203
- Brunner, M., Krauss, S. & Kunter, M (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence, 36*, 403–421. doi: 10.1016/j.intell.2007.11.002
- Buchwald, F., Fleischer, J. & Leutner, D. (2015). A field experimental study of analytical problem solving competence – Investigating effects of training and transfer. *Thinking Skills and Creativity, 18*, 18–31. doi: 10.1016/j.tsc.2015.04.009
- Buchwald, F., Fleischer, J., Rumann, S., Wirth, J. & Leutner, D. (2017). Training in components of problem-solving competence: An experimental study of aspects of the cognitive potential exploitation hypothesis. In D. Leutner, J. Fleischer, J. Grünkorn & E.

- Klieme (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 315–331). Berlin: Springer. doi: 10.1007/978-3-319-50030-0\_19
- Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research*, 36, 589–609. doi: 10.1207/S15327906MBR3604\_05
- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Cassady, J. C. & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27, 270–295. doi: 10.1006/ceps.2001.1094
- De Lange, J. (1996). Real problems with real world mathematics. In C. Alsina, J. M. Alvarez, M. Niss, A. Pérez, L. Rico & A. Stard (Eds.), *Proceedings of the 8th international congress on mathematical education, Sevilla July 1996* (pp. 83–110). Sevilla, Spain: Thales.
- Fleischer, J., Koeppen, K., Kenk, M., Klieme, E. & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Zeitschrift für Erziehungswissenschaft*, 16(Suppl. 1), 5–22. doi:10.1007/s11618-013-0379-z
- Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. N., Marchione, K. & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, 43, 1460–1473. doi: 10.1037/0012-1649.43.6.1460
- Fischer, A., Greiff, S., Wüstenberg, S., Fleischer, J., Buchwald, F. & Funke, J. (2015). Assessing analytic and interactive aspects of problem solving competency. *Learning and Individual Differences*, 39, 172–179. doi: 10.1016/j.lindif.2015.02.008
- Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 35–72). Hillsdale, NJ: Erlbaum.

- Gustafsson, J.-E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407–434. doi: 10.1207/s15327906mbr2804\_2
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Berlin: Springer. doi: 10.1007/978-3-642-20072-4\_7
- Heller, K. A. & Perleth, C. (2000). *KFT 4-12+R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Hogrefe.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistung. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule: Enzyklopädie der Psychologie, Serie Pädagogische Psychologie* (Bd. 3, S. 71–176). Göttingen: Hogrefe.
- Hickendorff, M., Van Putten, C. M., Verhelst, N. D. & Heiser, W. J. (2010). Individual differences in strategy use on division problems: Mental versus written computation. *Journal of Educational Psychology*, 102, 438–452. doi: 10.1037/a0018177
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan & J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.
- Jensen, A. R. & Weng, L.-J. (1994). What is a good g? *Intelligence*, 18, 231–258. doi: 10.1016/0160-2896(94)90029-9
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52, 876–903.
- Klieme, E., Hartig, J. & Wirth, J. (2005). Analytisches Problemlösen: Messansatz und Befunde zu Planungs- und Entscheidungsaufgaben. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 37–54). Wiesbaden: VS.
- Koeppen, K., Hartig, J., Klieme, E. & Leutner, D. (2008). Current issues in competence modelling and assessment. *Zeitschrift für Psychologie / Journal of Psychology*, 216, 61–73. doi: 10.1027/0044-3409.216.2.61

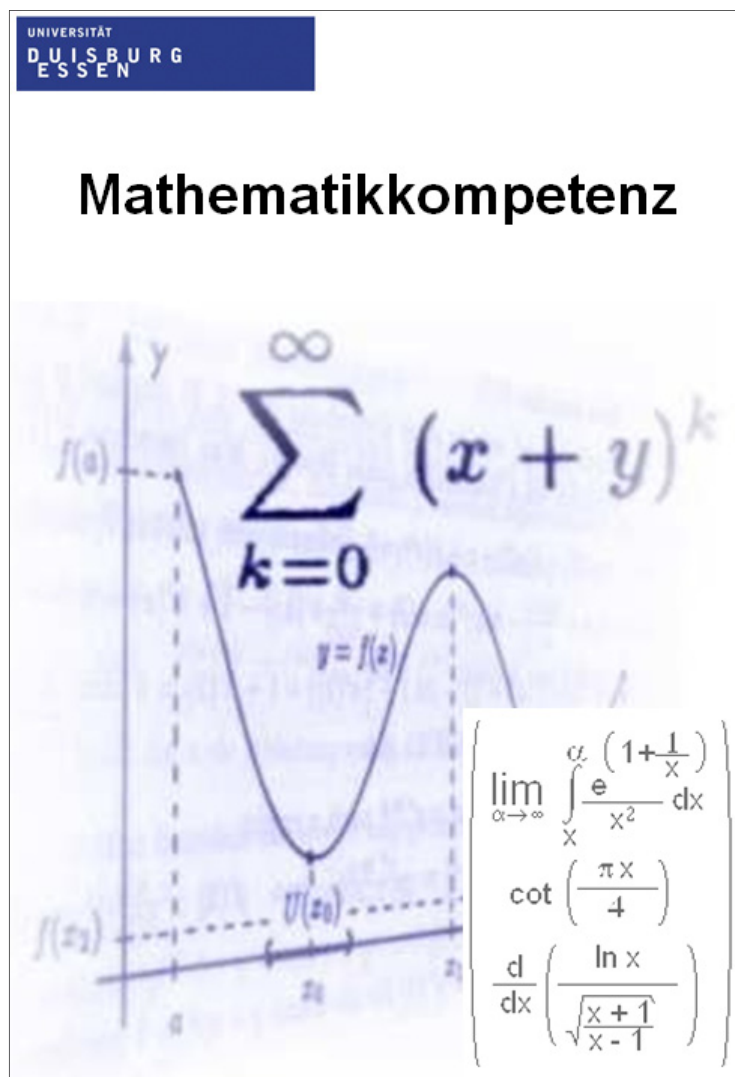
- Köller, O. & Baumert, J. (2008). Entwicklung schulischer Leistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6. Aufl., S. 735–768). Weinheim: Beltz.
- Kretzschmar, A., Neubert, J. C., Wüstenberg, S. & Greiff, S. (2016). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*, 54, 55–69. doi: 10.1016/j.intell.2015.11.004
- Leutner, D., Fleischer, J., Grünkorn, J & Klieme, E. (Eds.). (2017). *Competence assessment in education: Research, models and instruments*. Berlin: Springer. doi: 10.1007/978-3-319-50030-0
- Leutner, D., Fleischer, J., Wirth, J., Greiff, S. & Funke, J. (2012). Analytisches und dynamisches Problemlösen im Lichte internationaler Schulleistungsvergleichsstudien. *Psychologische Rundschau*, 63, 34–42. doi: 10.1026/0033-3042/a000108
- Leutner, D., Funke, J., Klieme, E. & Wirth, J. (2005). Problemlösefähigkeit als fächerübergreifende Kompetenz. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 11–19). Wiesbaden: VS. doi: 10.1007/978-3-322-85144-4\_2
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147–175). Münster: Waxmann.
- Leutner, D., Wirth, J., Klieme, E. & Funke, J. (2005). Ansätze zur Operationalisierung und deren Erprobung im Feldtest zu PISA 2000. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 21–36). Wiesbaden: VS.
- Lotz, C., Sparfeldt, J. R. & Greiff, S. (2016). Complex Problem Solving in Educational Contexts – Still something beyond a “good g”? *Intelligence*, 59, 127–138. doi: 10.1016/j.intell.2016.09.001

- McArdle, J. J. & Grimm, K. J. (2010). Five steps in latent curve and latent change score modeling with longitudinal data. In K. van Montfort, J. H. L. Oud & A. Satorra (Eds.), *Longitudinal research with latent variables* (pp. 245–273). Heidelberg: Springer. doi: 10.1007/978-3-642-11760-2\_8
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. doi: 10.1037/0003-066X.50.9.741
- OECD. (2004). *Problem solving for tomorrow's world. First measurements of cross-curricular competencies from PISA 2003*. Paris: OECD Publishing.
- OECD. (2005). *PISA 2003: Technical report*. Paris: OECD Publishing.
- Owens, M., Stevenson, J., Hadwin, J. A. & Norgate, R. (2012). Anxiety and depression in academic performance: An exploration of the mediating factors of worry and working memory. *School Psychology International*, 33, 433–449. doi: 10.1177/0143034311427433
- Rost, D. H. (2009). *Intelligenz Fakten und Mythen*. Weinheim: Beltz.
- Rost, J., Walter, O., Carstensen, C. H., Senkbeil, M. & Prenzel, M. (2004). Naturwissenschaftliche Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 111–146). Münster: Waxmann.
- Scherer, R. & Tiemann, R. (2014). Evidence on the effects of task interactivity and grade level on thinking skills involved in complex problem solving. *Thinking Skills and Creativity*, 11, 48–64. doi: 10.1016/j.tsc.2013.10.003
- Schmiedek, F & Wolff, J. K. (2010). Latente Wachstumskurvenmodelle. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 1017–1029). Wiesbaden: VS. doi: 10.1007/978-3-531-92038-2
- Stern, E. (2001). Intelligence, prior knowledge, and learning. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 7670–7674). Amsterdam: Elsevier. doi: 10.1016/B0-08-043076-7/02383-4

- Stern, E. (2003). Lernen ist der mächtigste Mechanismus der kognitiven Entwicklung: Der Erwerb mathematischer Kompetenzen. In W. Schneider & M. Knopf (Hrsg.), *Entwicklung, Lehren und Lernen – Zum Gedenken an Franz Emanuel Weinert* (S. 207–217). Göttingen: Hogrefe.
- van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics*, 25(2), 2–9.
- Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373–392). Thousand Oaks, CA: Sage.
- Zöttl, L., Heinze, A. & Reiss, K. (2007). Problemlösen im Kontext – Unterschiede in der Bearbeitung von Alltagsproblemen und mathematischen Problemen. In A. Peter-Koop & A. Bikner-Ahsbals (Hrsg.), *Mathematische Bildung – Mathematische Leistung. Festschrift für Michael Neubrand zum 60. Geburtstag* (S. 217–232). Hildesheim: Franzbecker.

## 7 ANHANG

### Anhang A1: Testheftcover und einführender Text in der Kontextbedingung „Mathematik“



Die Fähigkeit zum Umgang mit Mathematik ist in der heutigen Gesellschaft eine entscheidende Kompetenz, die zunehmend an Bedeutung gewinnt. So scheint eine hohe Mathematikkompetenz in vielen Bereichen auch für eine erfolgreiche Berufsausbildung oder ein erfolgreiches Studium wichtig zu sein.

## Anhang A2: Einführendes Item in der Kontextbedingung „Mathematik“

### **WOHNUNGSMIETE**

In einer Großstadt kostete 1985 eine 70 m<sup>2</sup>-Wohnung 1000 DM Miete pro Monat. Seit 1985 stieg der Mietpreis alle 5 Jahre um 20 %.

**Welche Monatsmiete musste dann 1995 für diese Wohnung gezahlt werden? Schreibe auf, wie du gerechnet hast.**



## Anhang B1: Testheftcover und einführender Text in der Kontextbedingung „Problemlösen“



Die Fähigkeit, Problemstellungen aus dem alltäglichen Leben zu verstehen und zu lösen, ist in der heutigen Gesellschaft eine entscheidende Schlüsselkompetenz, die zunehmend an Bedeutung gewinnt. So scheint eine hohe Problemlösekompetenz in vielen Bereichen auch für eine erfolgreiche Berufsausbildung oder ein erfolgreiches Studium wichtig zu sein.

## Anhang B2: Einführendes Item in der Kontextbedingung „Problemlösen“

### ENERGIEBEDARF

Bei dieser Aufgabe geht es um den empfohlenen täglichen Energiebedarf einer Person. Die folgende Tabelle stellt diesen Bedarf für verschiedene Personen in Kilojoule (kJ) dar.

#### Empfohlener täglicher Energiebedarf für Erwachsene

		<i>MÄNNER</i>	<i>FRAUEN</i>
Alter (Jahre)	Grad der Aktivität	Energiebedarf (kJ)	Energiebedarf (kJ)
Von 18 bis 29	Leicht	10660	8360
	Mittel	11080	8780
	Schwer	14420	9820
Von 30 bis 59	Leicht	10450	8570
	Mittel	12120	8990
	Schwer	14210	9790
60 und älter	Leicht	8780	7500
	Mittel	10240	7940
	Schwer	11910	8780

#### Grad der Aktivität nach Berufen

##### Leicht:

Verkäufer  
Büroarbeiter  
Hausfrau

##### Mittel:

Lehrer  
Verkäufer (Außendienst)  
Krankenschwester

##### Schwer:

Bauarbeiter  
Arbeiter  
Sportler

Herr David Eder ist ein 45-jähriger Lehrer.

**Wie hoch ist sein empfohlener täglicher Energiebedarf in kJ?**

Antwort:..... Kilojoules

## **Anhang C: Lebenslauf**

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten